Airway Basal Cells in Development, Injury-Repair, and Homeostasis

Ying Yang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Airway Basal Cells in Development, Injury-Repair, and Homeostasis

Ying Yang

Basal cells (BCs) are multipotent tissue-specific stem cells of a variety of organs including the skin, digestive and respiratory tract. BCs are broadly identified by expression of Krt5, Krt14 and the transcription factor p63. In the adult airways, BCs are not only important for normal maintenance but also crucial for epithelial repair after injury. However, the embryonic origin of these adult stem cells remains elusive. Previous reports showed that $p63^+$ cells appear early during airway development, but these do not express markers of adult BCs, raising the question whether these cells represent BC precursors. Moreover, little was known whether embryonic BCs have an impact in the adult pool of progenitors that mediate responses of the lung to injury or pulmonary diseases. The goal of this thesis is to address these gaps of knowledge using a variety of technologies, including functional and lineage tracing analysis *in vivo* in mouse genetic models, injury modeling, high-throughput profiling and gene regulation approaches.

This thesis is to comprehensively characterize airway BCs in development, injury-repair, and homeostasis. These studies revealed a previously unrecognized broader role of embryonic $p63^+$ cells in the establishment of the stem cell pools of the lung pre and postnatally. Surprisingly, lineage analysis showed that early in development these cells were able to generate all epithelial cell types of the airways and alveolar compartment. However, as development proceeds, they underwent two sequential lineage segregation events to finally generate two regionally distinct adult stem cell pools. One of these became the well-known BCs that populate extrapulmonary airways through an undescribed maturation process from the perinatal stage to adulthood, and the other was identified as a rare stem cell pool in the pseudostratified epithelium of intrapulmonary airways which maintained immature and quiescent throughout lifetime. Moreover, the latter responded uniquely to lung injury induced by H1N1 viral infection. Recent studies have demonstrated that BC-like $p63^+$ $Krt5^+$ cell clusters ("$Krt5^+$ pods") are ectopically present in the areas of severe alveolar injury by H1N1 viral infection. The presence of these pods has been associated with pathological scars in several human pulmonary diseases including idiopathic pulmonary fibrosis (IPF) and

acute respiratory distress syndrome (ARDS). However, their cellular origin has been intensely debated. This thesis showed that this rare progenitor pool is established during embryonic development when airways are still branching. Further characterization demonstrated a p63 gene dosage dependency in the specification/maintenance of this rare progenitor pool. By utilizing multiple lineage-tracing lines, an underappreciated diversity of this pool was revealed by showing a novel subpopulation carrying secretory lineage marker spatially restricted to intrapulmonary airways. Further molecular characterization and genetic manipulation of this rare progenitor pool may provide valuable cues to understand the pathogenesis about pulmonary disorders and to develop effective therapies.

Moreover, the molecular signatures of tracheal embryonic E18.5 preBCs and adult TrBCs were generated through high-throughput profiling, which provided hints about the genetic regulation of airway BC maturation process and generated potential molecular landmarks for the *in vitro* ES/iPS cell differentiation towards airway BCs. In addition, single cell RNA-sequencing analyses revealed heterogeneity of adult BCs in the tracheal and esophageal epithelia. Lastly, candidate master regulators of their differentiation programs in homeostatic and metaplastic states were identified through unbiased systems biology algorithms, which will be further validated in functional assays in the near future.

Taken together, the studies in this thesis comprehensively characterized airway BCs in development, injury-repair and homeostasis. This thesis work showed the newly identified $p63^+$ airway progenitors before E10.5 are multipotent for all lung epithelial lineages and this multipotency gets restricted to proximal fate at E10.5. In the adult injury-repair, this thesis work for the first time revealed that the H1N1-induced $Krt5^+$ pods are generated by bronchial $p63^+$ $Krt5^-$ progenitors, which originate from a subpopulation of E13.5 intrapulmonary $p63^+$ progenitors. At homeostasis, this thesis work uncovered a previously underappreciated heterogeneity of BCs in both airways and esophagus, and provided molecular foundations for further explorations into the mechanistic perspectives of BC cellular identity maintenance.

Table of Contents

List of Figures

Acknowledgments

I would like to thank my mentor Dr. Wellington V. Cardoso, for his support and appreciation throughout my five-year research life in his lab. Whenever I encountered problems, either in experiments or emotionally, Wellington's patience, enthusiasm and optimism would help me go through. Without his guidance, I could not finish my PhD study.

I would like to thank my thesis committee, Dr. Michael Shen, Dr. Hans Snoeck, Dr. Jianwen Que and Dr. Andrea Califano, and my qualifying committee, Dr. Cathy Mendelsohn and Dr. Frank Costantini, for their inspiring ideas, thoughtful comments and valuable time in shaping my projects and polishing my research.

I am also grateful to all of my collaborators, Dr. Jining Lu, Dr. Jianming Xu, Dr. Michael Schotsaert, Dr. Adolfo Garcia-Sastre, Dr. Iannis Aifantis, Dr. Maria Guillamot, Dr. Chao Lu, Dr. Xiao Chen, John McGuire, Dr. Hongxu Ding, for generously sharing mouse lines, valuable reagents, challenging techniques, computational analyses and friendship!

I thank my labmates, colleagues in CCHD, classmates and staff from the Department of Genetics and Development, for helping and supporting me in my research and study from the very start of my PhD journey.

A special gratitude goes to all of my friends, Tomato, Jingshu Huang, Xiaoqing Zhang, Wei Liu and others. Thank you all for making my PhD life colorful and wonderful!

Lastly, I would like to thank the mice sacrificed for my research. We were and will continue doing the right thing to explore the unknown and to benefit the world.

Dedication

This thesis is dedicated to my beloved parents, Yaojun Yang and Cuizhen Huang, who are always proud of me and supporting me to choose my life and to chase my dream. Also, to my best friend, Dr. Hongxu Ding, whose contagious passion and big smile inspire me to embrace the challenge with courage and determination.

Chapter 1. Introduction

Basal cells (BCs) are multipotent adult stem cells of a variety of tissues, including skin, esophagus, olfactory and airway epithelia, salivary and mammary glands. BCs are named for their proximity to the underlying basal lamina. They share common markers, including transcription factor p63 (tumor protein 63; transformation related protein 63), intermediate filaments Krt5 (cytokeratin 5), Krt14 (cytokeratin 14), and transmembrane integrin receptors Itga6 (CD49f), Itgb4 (CD104) and nerve growth factor receptor NGFR (p75)[1]. Multiple studies indicate that tissue-specific BCs serve as adult stem cells with capacities to self-renew and differentiate into other lineages of their tissue of origin, which underpins the importance of BCs in maintaining the integrity of distinct epithelia during homeostatic turnover and injury repair[2-5].

In the respiratory system, BCs account for about 6-30% in the pseudostratified epithelium in proximal airways depending on location. Bountiful data from *in vivo* lineage tracing (*Krt5-CreERT2 and Krt14-CreERT2*) and *in vitro* cell culture (air-liquid interface culture and matrigel-embedded 3D organoid culture) have presented evidence that BC can self-renew in long term and differentiate into luminal multiciliated and secretory cells, fitting the definition of adult stem cells[2,4,6]. The stem cell features of airway BCs can be hijacked during diseases or injury-repair. Excessive self-renewal leads to basal cell hyperplasia, and with additional triggers, may result in tumor formation. Inappropriate differentiation can lead to disruption of the luminal lineage balance with overrepresentation of goblet cells, causing excessive production of mucus and physical obstruction of small airways[7].

While there has been a wealth of studies to characterize signaling pathways regulating the regenerative activities of airway BCs in adulthood, the embryonic origins of these adult stem cells remain elusive. Previous reports showed that $p63^+$ cells appear early during airway development, however, these cells do not express canonical markers of adult BCs such as Krt5, NGFR[8-10]. Moreover, little was known whether embryonic $p63^+$ cells have similar potential as adult BCs, and how they contribute to the stem cell pool and the luminal compartment of airways in development, adulthood, and respond to severe injury or pulmonary diseases. The goal of this thesis is to address these gaps of knowledge using a variety of technologies, including functional and lineage tracing analysis *in vivo* in mouse genetics models, injury modeling, high-throughput profiling and gene regulation approaches.

1-1. Basic structure of the respiratory system

The mammalian respiratory system is a highly complex tree-like network of branching tubules and a large alveolar compartment specialized for gas exchange, designed to facilitate air breathing for terrestrial adaptation. The trachea bifurcates into two main stem bronchi, which further branch into multiple generations of airways connected with alveoli at the most distal ends. The cellular composition and 3D organization vary along proximal-distal axis, fulfilling the distinct functions in not only air transport and gas exchange, but also mucociliary clearance and innate immune defense demanded during the interaction between our body and external environment[11].

The conducting airways are comprised of trachea, bronchi and bronchioles. Large airways (trachea, bronchi are lined with a pseudostratified epithelium, with the architecture becoming progressively simplified into columnar epithelium towards the terminal bronchioles (small airways)[12]. The major cell types in conducting airways are secretory club and goblet cells, multiciliated cells, neuroendocrine cells, BCs and recently identified rare cell populations including ionocytes, tuft cells, and brush cells[13]. The distribution of each cell type also varies regionally to maximize the performance of particular function. BCs are widely accepted as adult stem cells which proliferate and generate luminal lineages in homeostasis and injury response. Multiciliated cells, are characterized by expression of the forkhead transcription factor Foxj1 and acetylated α-tubulin, are more abundant in proximal areas. These cells are terminally differentiated, and project hundreds of motile cilia towards the airway lumen to move external particles or pathogens through synchronized beating as part of the mechanism of mucociliary clearance[14,15]. Secretory cells comprise a heterogeneous cell population which have abundant secretory granules, and express multiple secretoglobins, including Scgb1a1 (CC10/CCSP), Scgb3a2, Scgb3a1, mucins (Muc5ac, Muc5b) and play critical roles in oxidative stress reduction, innate immune response, and xenobiotic metabolism[16,17,18]. Pulmonary neuroendocrine cells sense the environmental cues and respond by orchestrating neural system, immune cells, and pulmonary epithelial cells through neuropeptide secretion[19,20]. They are marked by expression of proneural bHLH transcription factor Ascl1 (Mash1), Cgrp (Calcitonin), Pgp9.5 (Ubiquitin carboxy-terminal hydrolase L1, Uchl1) and others. Neuroendocrine cells generally form clusters (NEB,

neuroendocrine body, innervated by both afferent and efferent neurons) at the branchpoint of each airway generation. Others remain isolated and intermixed with multiciliated and secretory cells.

In alveolar region, two major epithelial cell types are present, alveolar type 1 cell (AT1), and alveolar type 2 cell (AT2). AT1 cells, marked by Pdpn (T1α), have extremely large size and cover more than 95% of the gas exchange surface. They undergo a non-proliferative two-step process, flattening and folding during development, to form remarkably thin cellular structure juxtaposing the alveolar capillary network[21]. AT2 cells are cuboidal, intercalated with large AT1 cells. They produce surfactants, including surfactant-associated protein C (Spc, Sftpc), and lipids to reduce the surface tension of alveoli.

The respiratory systems in mice and human are largely conserved , but differ in their regional cellular composition. In humans, BCs are found throughout intrapulmonary airways, extending to terminal bronchioles. In mice, BCs are restricted to extrapulmonary airways including trachea and main stem bronchi, and are largely devoid in intrapulmonary airways. It is widely accepted that the mouse trachea can be used as a model to study human airway biology[1,11]. In addition, goblet cells are almost absent from the adult mouse airways. These differences between mouse and human respiratory system may underlie the changes in living environment and evolutionary adaptation.

1-2. Development of the mouse respiratory system

The mouse respiratory domain is established at E9.0, marked by induction of the homeodomain transcription factor Nkx2-1 in the ventral foregut endoderm. The patterning of foregut endoderm by BMP antagonist Noggin from dorsal notochord and BMP/FGF/WNT from ventral cardiac mesoderm specifies the ventral trachea and dorsal esophagus. Genetic manipulation of BMP signaling pathway affects this dorsal-ventral patterning, as shown by foregut ventralization in Noggin mutants[22] and dorsalization in Bmpr1a/b mutants[23]. Early studies using foregut culture suggested that different thresholds of FGF signaling from the cardiac mesoderm are critical for the specification of respiratory domain in the foregut patterning process[24]. The essential role of WNT signaling in this process is supported by the complete lung agenesis with no Nkx2-1 expression in the middle foregut of Wnt2/2b null mice. Moreover, activation of canonical WNT/β-catenin signaling leads to ectopic lung endoderm fate in esophagus and stomach endoderm[25,26]. Additional

signaling pathways including SHH (Sonic hedgehog) and RA (Retinoic acid) have been suggested to be integrated with WNT, BMP and FGF in specification of the respiratory domain[27-29].

Currently, the only known existing marker of lung identity is Nkx2-1[30], which is not only expressed in lung, but also in thyroid and brain. An antagonized interaction between ventral Nkx2-1 and dorsal Sox2 has been revealed to be critical in the specification/maintenance of tracheal versus esophageal fates[10,31].

From E9.5, the ventral trachea and dorsal esophagus starts to separate, and this process is completed by E11.5, splitting the initial single-lumen foregut into two tubes[32]. The detailed mechanisms regulating this separation morphogenesis is still in debate. A “splitting and extension model” has been proposed based on live imaging of foregut cultures. The separation initiates from the site of lung bud outgrowth with a saddle-like structure moving anteriorly and a nascent trachea and esophagus extending posteriorly[33]. Alternative models include the fusion of endodermal ridges in the midline[32] and extension of a tracheoesophageal septum by formation of a mesenchymal condensation[34]. Failure of this separation leads to esophageal atresia with/without tracheoesophageal fistula (EA/TEF), which is a relatively common birth defect with high morbidity[35].

By E9.5, two primary lung buds evaginate from the ventral wall of foregut. This primary budding requires Fgf10-Fgfr2 signaling, as epithelial deletion of Fgfr2 phenocopies the mesenchymal disruption of Fgf10 and results in lung agenesis[36-38]. Secondary buds form at E10.5, with the first appearance of Spc expression in the tip region[39], and the establishment of the lobation pattern of the right and left lungs. From E10.5, epithelial tubules undergo reiterative growth and budding, referred as branching morphogenesis, until approximately E16.5 to form the bronchial tree. Coupled with branching morphogenesis is proximal-distal patterning, which is crucial to specify the distal future gas-exchange domain and the proximal airway domain. The distal progenitors express transcription factor Sox9 (sex determining region Y-box 9) and Id2 (inhibitor of DNA binding 2). Proximal progenitors express Sox2 and further differentiate into distinct airway lineages including secretory, multiciliated cells and others. Lineage tracing analyses using *Id2-CreERT2* and *Sox9-CreERT2* provided evidence supporting that the epithelial progenitors in the distal compartment self-renew extensively and continuously generate proximal airway descendants in a centripetal fashion during branching morphogenesis. By E16.5, these distal progenitors no longer generate proximal cells, but instead become restricted to distal fate by differentiating further into AT1 and AT2 lineages[40,41]. Signaling

pathways such as BMP, WNT are shown to be crucial in fate specification or maintenance of distal progenitors[42,43]. The fate switch from distal progenitors to proximal airway progenitors is still poorly understood. Hippo/Yap pathway has been suggested to play key roles in this process. At the transition zone between $Sox9^+$ distal domain and $Sox2^+$ proximal domain, Yap undergoes a nucleocytoplasmic shift, and responds to proximalizing local cues, which ultimately leads to Sox2 expression and airway progenitor cell specification[44,45].

Still during the formation of branchial tree, airway progenitors which initiate differentiation program towards multiple airway epithelial cell lineages. By ~E12.5 to E14.5, progressive acquisition of secretory, multiciliated and neuroendocrine cell fate starts, marked by expression of Scgb3a2, Foxj1 and Ascl1, respectively. Notch signaling plays significant roles in this cell fate selection, and disruption of Notch leads to overrepresentation of multiciliated cells and neuroendocrine cells at the expense of secretory cells[46-48].

A single-cell RNA-sequencing analysis suggested that a bipotential progenitor population exists in the distal airways at E16.5 to form alveolar AT1 and AT2 lineages. These bipotential progenitors co-express markers for AT1 and AT2, and progressively branch into either AT1 or AT2 lineages from E16.5 to E18.5[49].

1-3. Current understanding of airway BCs

Airway BCs, marked by p63, Krt5, and Krt14 (~20% of mouse airway BCs express Krt14 in homeostasis), are considered as the major adult stem cells in the pseudostratified epithelium of mouse trachea. They are named for their proximity to basal lamina, and consist of ~20-30% of the total epithelium. Previous lineage tracing analyses using *Krt14-CreER* and *Krt5-CreER* (transgene) suggested that airway BCs self-renew and contribute to maintenance of luminal lineages (multiciliated and secretory cells)[2,50]. Isolated airway BCs from both mice and human can expand *in vitro* and differentiate into multiciliated and secretory cells in air-liquid interface culture or 3D matrigel-embedded organoid culture[2,51], further supporting the adult stem cell identity of airway BCs. Additionally, early xenograft studies demonstrated that rat tracheal BCs can reconstitute the denuded tracheas and give rise to fully differentiated mucociliary epithelium afterwards[52]. In response to injury by SO2, luminal cells die and detach from basement membrane, while BCs expand to cover the wound and quickly regenerate the whole epithelium in two weeks[2,53]. Immediately following injury, a functional segregation of airway BCs occurs. A subpopulation of airway BCs begin to express c-myb,

which further differentiate into multiciliated cells, while others express N2ICD, generating secretory cells following[54]. The importance of Notch signaling in this regenerative process has also been elucidated earlier using both *in vivo* genetic manipulation and *in vitro* organoid culture[3]. Additional regulators including IL-6/STAT3 and Grhl2 (transcription factor Grainyhead-like 2) have been identified as key factors controlling cell fate selection from BCs to multiciliated versus secretory lineages[53,55].

The potential heterogeneity of airway BCs has been a topic of intense investigation. Long-term clonal analysis and mathematical modeling suggested that airway BCs comprise equal numbers of multipotent stem cells and committed precursors. 94% of basal stem cells undergo asymmetric division to produce one stem cell and one basal luminal cell (this population slightly upregulates Krt8, but still retain p63 and Krt5 expression) in a hypothesized process called stochastic homeostasis. The basal luminal precursor further differentiates into a secretory descendant over the course of ~11 days. The secretory descendant then commits to differentiation towards multiciliated fate. Direct differentiation from BCs to multiciliated cells was considered as a negligible event[4]. Krt14 was proposed to be a marker of a unipotent stem cell population[6], but this was not supported by the clonal analysis.

The regulation of airway BC activities in homeostasis is important, as the stem cell features of this population can be hijacked during injury-repair and tumor initiation. *In vivo* genetic manipulation and *in vitro* culture studies revealed multiple signaling pathways as regulators of airway BC proliferation and differentiation. EGFR (epidermal growth factor receptor) activity, canonical WNT/β-catenin and YAP pathway serve as agonist to promote proliferation of airway BCs[56-59], while Notch plays an antagonist role in BC self-renewal[3]. FGF signaling shows distinct effects through different receptors. Conditional deletion of Fgfr1 using *Krt5-CreER* driver enhanced airway BC proliferation[60]; in contrast, Fgfr2 deletion in Krt5$^+$ adult BCs led to defect in self-renewal and differentiation, partially through regulating the key transcription factor Sox2[61]. Furthermore, BMP and TGFβ signaling activity is suppressed in adult airway BCs in homeostasis, and this signaling inhibition has been exploited in *in vitro* cell culture to expand BCs with differentiation suppression[62].

In contrast to the extensive investigations on adult airway BC activities, not much was known about the BC ontogeny in part due to lack of genetic tools. The master regulator of BCs, p63, has been reported to start expression in ventral tracheal domain from E10.5[9,10]. Another canonical BC marker Krt5 has been

reported to start expression in airway epithelium from E15.5. Even though by E18.5 p63/Krt5 double-positive BC-like cells were identified in trachea, these cells do not express additional BC markers such as NGFR until after birth[8]. It is unclear whether these cells are direct precursors for adult BCs and whether they possess similar potentials as their adult counterparts. An *in vitro* colony expansion assay was performed, attempting to investigate the ontogeny of airway BCs. In this study, ITGB4^{hi} preBCs were separated from the rest Nkx2-1^{+} epithelial cells, but no difference in colony-forming efficiency and mucociliary differentiation was identified[9], ruling out the possibility that ITGB4^{hi} preBCs were the only precursors for airway BCs. Overexpression of Fgf10 using R26-rtTA; Tet-Fgf10 after E12.5 leads to ectopic p63^{+} cells while Fgf10 mutants exhibit decreased p63^{+} cell number in lower trachea, underlying the importance of Fgf10 signaling in airway BC development[63]. Epigenetic remodeling also plays essential roles: Genetic deletion of Ezh2 in Shh-Cre; Ezh2 f/f animals results in advanced maturation of tracheal BCs (expression of Krt5 as early as E14.5) and ectopic appearance of p63^{+} Krt5^{+} cells in intrapulmonary airways, and this effect is partially explained by the repression of Igf1 (Insulin growth factor 1)[64,65]. Thus, it remains unclear when and how airway BCs are specified and how the embryonic progenitors contribute to this stem cell pool and the luminal compartment of airways in development.

1-4. p63, the master regulator of BC identity in common

p63 is a highly conserved homolog of p53 transcription factor family. It is predominantly expressed in BCs in the epithelia of multiple tissues, including epidermis, esophagus, olfactory epithelium and airway epithelium[5,66,67]. p63 is a regulator of BC identity, and it has been suggested to play essential roles in epithelial morphogenesis and maintaining stemness in progenitor cells by multiple loss-of-function studies[5,68-70].

p63 gene is located on chromosome 3 in human and chromosome 16 in mice. p63 is expressed from two promoters, producing two major isoforms TAp63 and ΔNp63. The full-length TAp63 contains a transactivating domain in the N-terminal, while ΔNp63 uses an alternative promoter in an additional exon 3, resulting in a N-terminal truncated transcript version without the transactivating domain. TAp63 isoform is thought to be restricted primarily in the oocytes[71], however recent studies uncovered that TAp63 may be

expressed to exert functions in skin development, ciliogenesis and early embryogenesis[72-75]. ΔNp63 is widely expressed in epithelial BCs, showing undoubted significance in epithelial morphogenesis.

p63 null animals die immediately after birth, due to dehydration from defective skin development. Striking defects were also observed in limb, tail, craniofacial, and genital development[68,69,76,77]. Detailed examination in *in vitro* clonal assays and *in vivo* histology analyses revealed that p63 is essential to maintain the proliferative activity of epithelial stem cells in thymus and epidermis[71]. Additionally, whether p63 is essential for the epithelial stratification remains controversial, as conflicted results were presented from different groups[71,77,78]. Recent studies using hESC differentiation platform and genome-wide epigenetic mapping suggested that p63 is not required for the specification of epidermal lineage, but essential for the maturation process afterwards to functional keratinocytes[79,80]. Other than epidermis, the self-renewing capacity of mammary stem cells is also controlled by p63, specifically, ΔNp63, as loss of one functional allele of ΔNp63 dramatically reduces the fat pad reconstitution efficiency and frequency of mammary stem cells[81].

In the airway epithelium, ΔNp63 is critical for the formation of BCs. Tracheal epithelium in p63 null animals exhibits simplified columnar architecture with no BCs instead of the pseudostratified structure. Intriguingly, the airway epithelium was found over populated by multiciliated cells at the expense of secretory cells when analyzed by immunohistochemistry for Tub4, a marker of motile cilia[82]. Conditional deletion of p63 from adult airway BCs using *Krt5-CreER* driver caused spontaneous differentiation of p63 null cells towards luminal lineages, suggesting an essential role of p63 in maintenance of BC identity[59]. A lentiviral silencing of p63 in human airway BC culture led to decrease in proliferation and increase in cellular senescence[83].

1-5. Regionally distinct stem cell populations and injury-induced epithelial plasticity

The lung epithelium is highly quiescent compared to other organs with high turn-over rate like esophagus, intestine, skin and hematopoietic system. The direct exposure to environmental agents and pathogens renders the lung sensitive for damage and calls for regenerative capacity. As epithelial cells are relatively immobile, and the cellular composition varies dramatically in distinct anatomic locations, different regionally-specific stem cells are proposed to be responsible for the homeostatic maintenance and injury repair along proximal-distal axis.

In the pseudostratified epithelium of extrapulmonary airways, BCs continuously self-renew and give rise to luminal multiciliated and secretory descendants[2]. In mouse intrapulmonary airways, BCs are largely devoid, and Scgb1a1$^+$ secretory cells self-renew in long term and give rise to multiciliated cells in homeostasis and in injury such as naphthalene which kills the majority of the secretory cells[84]. Scgb1a1$^+$ secretory cells do not contribute to the regeneration of alveolar lineages[84]. A subpopulation of secretory cell, preferentially identified as juxtaposed to NEBs and near the junction of terminal bronchioles with alveoli, is relatively resistant to naphthalene injury due to lack of Cyp2f2 (cytochrome p450). These cells express high levels of Upk3a, Scgb3a2 and low levels of CC10. They can repopulate the intrapulmonary airways after naphthalene injury and contribute slightly to alveolar regeneration after bleomycin injury[16,85].

In the alveolar compartment, AT2 cells are considered as adult stem cells. Lineage tracing analyses using *Spc-CreER* provided good evidence that AT2 cells self-renew in long term and give rise to AT1 cells in homeostasis[86,87]. In targeted AT2 injury by genetic expression of DTA (diphtheria toxin A) under *Spc-CreER* driver, the surviving AT2 clonally expanded and generate both AT2 and AT1 cells[86]. AT2 cells are heterogeneous in WNT signaling activity. Two recent studies suggested that Axin2$^+$ AT2 cells, a WNT signaling-active AT2 subpopulation, serve as the stem cells for the whole alveolar epithelial compartment[88,89]. The Axin2$^-$ AT2 cells can convert to Axin2$^+$ progenitors in response to hyperoxia injury, as an autocrine secretion of Wnt2 from AT2 cells was induced, leaving the WNT-producing stromal niche dispensable[88]. However, in the accompanying report, this conversion was not observed, and Axin2$^+$ AT2 cells were solely responsible for alveolar regeneration after H1N1-induced acute lung injury[89].

Even though different regions in respiratory system can be maintained by the local resident stem cells which are considered more undifferentiated, reserved/facultative stem cells have been identified in different injury scenarios. Epithelial cells have under-appreciated plasticity. Differentiated cells might be recruited and fulfill the stem cell roles in specific injury challenges. For example, in extreme scenario that tracheal BCs were eliminated through genetic manipulation (Krt5-rtTA; tetO-DTA), secretory club cells were found to dedifferentiate into basal stem cell lineage and function as their endogenous counterparts in epithelial regeneration[90]. In addition, after partial pneumonectomy, unexpected conversion was considerably observed from terminally differentiated AT1 cells to AT2 cells, and this phenotypic switch from AT1 to AT2 can also be recapitulated in *in vitro* 3D organoid cultures[91]. Pulmonary neuroendocrine cells were shown

to generate secretory and multiciliated cells in *CGRP-CreER* lineage tracing after naphthalene injury[92]. Such plasticity is a universal phenomenon adopted in multiple epithelia, best studied in intestine. In intestine, Lgr5-expressing CBCs (crypt base columnar cells) continuously self-renew and generate secretory cells and absorptive enterocytes in homeostasis[93]. But when CBCs are genetically ablated, $Bmi1^+$ cells located at the +4 position of each crypt can dedifferentiate into CBCs to restore homeostasis[94]. Irradiation can even induce acquisition of stem cell features in the Paneth cells through activation of Notch signaling[95].

Another event in epithelial plasticity is the recruitment of stem cells from neighboring tissues or domains. In the respiratory system, this phenomenon was illustrated by the fate interchange between submucosal gland (SMG) stem cells and surface airway epithelial (SAE) stem cells. SMGs are small epithelial appendages composed of ducts and acini embedded in the interstitial layer of the upper conducting airways. The stem cell population in SMGs, Acta2 (α-SMA)$^+$ $Krt5^+$ $Krt14^+$ basal-like myoepithelial cells, serves as reserve stem cells for the SAE, recruited from the SMG microenvironment whenever necessary to regenerate both basal and luminal compartments in the SAE[96-98]. The descendants derived from SMG myoepithelial cells migrate into injured SAE, and gradually adopt airway BC features, presumably through microenvironmental reprogramming. Such microenvironmental reprogramming has been reported in thymic epithelial cells that adopt a hair follicle progenitor fate following epidermal transplantation[99]. The recruitment of stem cells from an adjacent region to repair a damaged epithelium depleted of its own stem cell pool has also been reported in interfollicular and hair follicle stem cells[100].

An intriguing acute lung injury model using H1N1 influenza virus infection was initially thought to be another example of this stem cell recruitment phenomenon. The H1N1 influenza virus induces wide-spread damage of both conducting airways and the alveolar compartment. Catastrophic epithelial destruction, as well as massive infiltration of immune cells, could be detected during 4dpi (days post injury) and 15dpi. The mice suffer from extreme weight loss from 4dpi but can gradually recover from 11dpi onwards. In the initial study, the lungs were thought fully recovered without fibrosis, eliciting the intriguing hypothesis that a novel adult stem cell population exists inside the lung with marvel regenerative potential. Since the characterization of this injury model in 2011, there have been heated debates, as a large number of p63-expressing $Krt5^+$ basal-like cells were found ectopically in the highly damaged lung parenchyma at 11dpi.

These p63$^+$ Krt5$^+$ cells formed discrete clusters, referred as Krt5$^+$ pods, were proposed to serve as stem cells, responsible for the regeneration of damaged airways and alveoli[101].

Extensive lineage-tracing studies have been performed in the past few years, aiming at the cellular origin of these Krt5$^+$ pods. DASCs (distal airway stem cells), double-positive for p63 and Krt5, were found in peribronchiolar region before injury. Lineage tracing using a *Krt5-CreERT2* transgenic line labeled this population, and showed robust labeling in the Krt5$^+$ pods post injury[102]. Another group using *Scgb1a1-CreERT2* lineage tracing line suggested the Krt5$^+$ pods might be derived from secretory cells and these pods further regenerate AT1 and AT2 cells[103,104]. However, these lineage tracing strategies suffered from a caveat of tamoxifen persistence, as tamoxifen administrated through intraperitoneal injection may not be cleared within 3 weeks or even as long as 1 month. Additionally, the *Scgb1a1-CreERT2* line can label a considerable AT2 population in homeostasis[84]. The marker used for describing AT1 regeneration, Pdpn, is shared by airway BCs, leading to the misunderstanding of alveolar regeneration in the initial studies. Chapman group using multiple CreERT2 knock-in lines for stringent lineage tracing analysis, found minimal contribution to Krt5$^+$ pods from either Spc$^+$ AT2 cells or CC10$^+$ secretory cells, and 13% of Krt5$^+$ pods bearing the *Krt5-CreERT2* lineage label. The putative migration of extrapulmonary airway BCs into the injured lung was excluded by their lung transplantation experiments. LNEPs (lineage-negative epithelial progenitors, defined as EpCAM$^+$ CC10-lineage$^-$ CD104$^+$ CD200$^+$ CD14$^+$ isolated from distal airways) were proposed as the major source for Krt5$^+$ pods based on the cell transplantation assay[105]. The potential contribution from AT1 lineage was further excluded in *Hopx-CreERT2* lineage tracing. But *Sox2-CreERT2* labeled robustly the ectopic Krt5$^+$ cells after injury, restricting the cellular origin of Krt5$^+$ pods in proximal airways[106]. Until recently, a follow-up study from Chapman group showed that the Krt5$^+$ pods were derived from a resident p63 lineage positive population in distal bronchioles. The Krt5$^+$ pods persisted in long term and did not result in regeneration of any alveolar structures, unless Notch or WNT signaling pathways were genetically manipulated[107]. The persistence of ectopic Krt5$^+$ pods was also associated with pathological scars identified in human pulmonary disorders including ARDS (acute respiratory distress syndrome) and IPF (idiopathic pulmonary fibrosis), further confirming the significance of deep understanding of the biology of these cells.

Chapter 2. Methods & Materials

2-1. Mice

p63-CreERT2 mice were generated and characterized as described in Lee et al. 2014[108]. This mouse line was originally in FVB background, with the CreERT2 coding sequence knocked into exon 4 which is shared by both deltaNp63 (major isoform in airway system) and TAp63 transcripts. *Sox9-CreERT2*, *Spc-CreERT2*, *Upk3A-CreERT2*, *N3-CreERT2* knock-in lines have been previously reported and were currently used for lineage tracing of embryonic distal lung epithelial progenitors (Sox9[109]), adult alveolar type II cells (Spc[87]), adult secretory cells associated with neuroendocrine bodies and terminal bronchioles (Upk3A[85]), Notch3-expressing mesenchymal and rare scattered epithelial cells (N3[110]), respectively. The *CC10-CreERT2* knock-in line was purchased from the Jackson Laboratory (B6N.129S6(Cg)-Scgb1a1tm1(cre/ERT)Blh/J)[84], and was currently used to follow the fate of adult club and Spc^+ $CC10^+$ bronchioalveolar progenitor cells. All knock-in lines were bred into *R26-tdTomato* (Jackson Laboratory, B6.Cg-Gt(Rosa)26Sortm14(CAG-tdTomato)Hze/J) for lineage-tracing analyses. All studies were approved by Columbia University Institutional Animal Care and Use committees (IACUC).

For time-pregnancy experiments, male mice from CreERT2 driver lines were mated with female mice from the *R26-tdTomato* reporter line. Noon of the day when the vaginal plug was identified was determined as E0.5. The developmental stages of TM administration for each lineage-tracing experiment were detailed below.

For adult lineage-tracing experiments, 6-8 week-old mice were used and both male and female were included. Details about TM administration and H1N1 infection were described below.

2-2. Tamoxifen Induction of Cell-Lineage Tracing

2-2a. Lineage Tracing in Embryonic Development

Tamoxifen (TM) was dissolved in sunflower seed oil (Sigma, T5648 and S5007) and administered by gavage to pregnant mice. To allow survival and analyses of postnatal stages, E18.5 embryos were transferred to adult CD1 foster mothers.

To determine the efficiency of recombination at different TM doses and developmental stages we treated *p63-CreERT2; R26-tdTomato* mothers with TM by oral gavage at gestation days 13.5 (160 mg/g body weight), 14.5 (80 mg/g), 17.5 (70 mg/g) and analyzed lungs after 24hr. IF and quantitative analysis showed efficient recombination with the percentage of lineage-labeled $p63^+$ tracheal cells averaging 35.7% ± 4.5%, 54.3% ± 1.6%, and 95.6% ± 1%, respectively (post hoc Tukey's test: E13.5 vs. E14.5 **adjusted P<0.01; E14.5 vs. E17.5 ***adjusted P<0.001; E13.5 vs. E17.5 ****adjusted P<0.0001.). Higher recombination efficiency was achieved at later developmental stages in spite of the lower TM doses, consistent with the overall increase in p63 expression levels perinatally.

TM-independent recombination in mice treated with vehicle at E13.5 and examined at E18.5 and postnatal P21 averaged <0.1% and 1.3% ± 0.5 of the $p63^+tdTomato^+$ tracheal cells, respectively.

To test for potential TM residual effects, we treated *p63-CreERT2; R26-tdTomato* E13.5 mothers with decreasing TM doses (160 µg/g, 10 µg/g and 4 µg/g body weight), looking for changes in the proportion of lineage-labeled cells in the basal versus luminal compartments at E18.5. Analysis of E14.5 tracheas confirmed the major dose-dependent decrease in recombination efficiency in $p63^+$ cells, from 35.7% to 6.3% and 0.6%, respectively. By E18.5 the significantly higher proportions of $p63^+$ $tdTom^+$ in total $tdTom^+$ cells were found at lower doses (43%, 56%, 57%, respectively; post hoc Tukey's test: 160 µg/g vs. 10 µg/g **adjusted P<0.01; 160 µg/g vs. 4 µg/g **adjusted P<0.01; 10 µg/g vs. 4 µg/g non-significant P>0.05. This scenario would be unlikely in the presence of sustained recombination and thus argued against a TM residual effect.

For labeling $p63^+$ cells before E14.5, we treated *p63-CreERT2; R26-tdTomato* mothers with 160 µg/g Tm by oral gavage (TM exposure at E8.5, E9.5, E10.5, E12.5 or E13.5).

To label $Sox9^+$ multipotent lung tip cells, we exposed *Sox9-CreERT2; R26-tdTomato* mothers with 160 µg/g TM by oral gavage at E9.5 or E11.5. Lungs were examined at E15.5 or E17.5.

2-2b. Lineage Tracing in Adult for H1N1 Injury Response

For lineage tracing in adult animals in H1N1-infected lungs, 6-8 -week-old *p63-CreERT2, CC10-CreERT2*, *p63+/+*, *Spc-CreERT2*, *Upk3A-CreERT2*, *N3-CreERT2* (male and female, all with *R26-tdTomato* reporter allele) mice were given 240 µg/g body weight TM in 5 sequential days via oral gavage. A chase period of 3 weeks was used to prevent artefactual tamoxifen residual activity before viral infection.

2-2c. Lineage Tracing in Adult for BC isolation

To efficiently and specifically label BCs in both trachea and esophagus in adult animals, 6-12 -week-old *p63-CreERT2; R26-tdTomato* mice were give 240 µg/g body weight TM once via oral gavage. A short chase period of 2-3 days was used to ensure accumulation of tdTomato in BCs and to minimize tdTomato labeling in luminal descendants.

2-3. Histology and Immunostaining

Embryonic and neonatal lungs were fixed in 4% paraformaldehyde in PBS at 4°C for 1 hour (earlier than E14.5), 4 hours (E14.5, E15.5), overnight (E18.5 and older; whole embryos). Adult lungs were inflated with 4% paraformaldehyde (25cm water column pressure) through trachea and fixed overnight. Samples were processed for frozen or paraffin-embedding.

Immunofluorescence (IF) was performed in tissue sections (6-8mm) blocked with 10% horse serum and 0.3% TritonX-100 (Sigma) for 1 hour at room temperature (rt). Primary antibodies were incubated in 1% bovine serum albumin (Sigma) and 0.3% TritonX-100 at 4°C overnight or 2 hours at rt. Sections were then washed with PBS and incubated with Alexa Fluor-conjugated secondary antibodies (1:500) and NucBlue Live Cell ReadyProbes Reagent (DAPI) (Life Technology) for 1 hr. After washing, samples were mounted with ProLong Gold antifade reagent (Life Technology). When necessary, antigen unmasking was done using Citric Based solution (Vector Labs H-3300) heated in microwave. Mouse primary antibody staining was done using M.O.M kit (Vector Labs BMK-2202). Whole-mount IF was performed using same reagents with elongated antibody incubation time. Confocal Microscopy was performed using a Zeiss LSM 710 confocal microscope.

For visualization of labeled antibodies in immunohistochemistry (IHC), Mouse on Mouse Elite Peroxidase kit (Vector Labs, PK-2200) and Vectastain Elite ABC-HRP kit (Vector Labs, PK-6100) were used. To block endogenous enzyme activity, 0.3% hydrogen peroxide in 0.3% horse serum was used. Following DAB staining, Alcian Blue staining was performed. Images were acquired on a Nikon Labophot 2 microscope equipped with a Nikon Digital sight DS-Ri1 charge-coupled device camera.

The following primary antibodies were used: rabbit anti-p63a (1:400, CST, 13109); chicken anti-Krt5 (1:500, Biolegend, 905901); rabbit anti-Krt5 (1:500, Biolegend, 905501); chicken anti-GFP (1:1000, Abcam,

ab13970); chicken anti-Krt8 (1:500, Abcam, ab107115); goat anti-CC10 (1:150, Santa Cruz, sc-9772); mouse anti-Ki67 (1:300, BD Biosciences, 550609); mouse anti-Ecad (1:100, BD Biosciences, 610181); mouse anti-Foxj1 (1:100, eBioscience, 14-9965); mouse anti-acetylated a-tubulin (1:2000, Sigma, T7451); rabbit anti-acetylated a-tubulin (1:500, CST, 5335); rat anti-Scgb3a2 (1:100, R&D, MAB3465); rabbit anti-Cgrp (1:1000, Sigma-Aldrich, C8198); rabbit anti-Sox9 (1:500, Millipore, AB5535); goat anti-Sox9 (1:200, R&D, AF3075); rat anti-Sox2 (1:200, eBioscience, 14-9811-82); hamster anti-Pdpn (1:50, Developmental Studies Hybridoma Bank); rabbit anti-proSpc (1:1000, Seven Hills, WRAB9337); mouse anti-SSEA1 (1:50, Santa Cruz, sc21702); rabbit anti-Nkx2.1 (1:100, Abcam, ab76013); rabbit anti-Notch3 (1:50, CST, 5276).

The following secondary antibodies were used: donkey anti-rabbit (conjugated with Alexa Fluor 488, 568, 647); donkey anti-chicken (conjugated with Alexa Fluor 488); goat anti-chicken (conjugated with Alexa Fluor 488, 647); donkey anti-goat (conjugated with Alexa Fluor 488, 568, 647); donkey anti-mouse (conjugated with Alexa Fluor 488, 568, 647); donkey anti-rat (conjugated with Alexa Fluor 488, 647); goat anti-hamster (conjugated with Alexa Fluor 647). All secondary antibodies were purchased from Thermo Fisher Scientific or Jackson ImmunoReseach.

2-4. EdU Incorporation

Analyses of cell proliferation in lineage-labeled cells was performed by exposing p63-CreERT2 pregnant mice to tamoxifen at designated stages through oral gavage and simultaneous intraperitoneal (i.p.) injection of EdU solution (1mg, 5mg/ml stock solution in PBS). EdU incorporation was assessed 24hr later in frozen tissue sections using the Click-iT EdU Alexa Fluor 647 Imaging kit (Thermo Fisher, C10340). Sections were then double-labeled with antibodies against selected cell markers using immunofluorescence.

2-5. Quantitative Real-time PCR

Whole tracheas were isolated (below the cricoid cartilage to carina) from *p63-CreERT2/p63-CreERT2* KO and *p63+/+* WT animals at E18.5. RNA was extracted using QIAcube and QIAGEN RNeasy Mini Kit. cDNA was synthesized using the SuperScript IV First-Strand synthesis system (Thermo Fisher). Gene expression was assessed using TaqMan Fast Universal PCR Master Mix (Applied Biosystems) and analyzed on a Step-One Plus instrument (Applied Biosystems). At least 4 tracheas were analyzed for each group. qRT-

PCR in whole tracheal homogenates ensured that conclusions from analysis of tissue sections were not influenced by sampling biases due to regional distribution of these cell types.

2-6. Fluorescence Intensity Profile Analysis

For E14.5, 150 cells from epithelium were randomly picked, circled around nuclear shape (DAPI). Fluorescent intensity and area (mm^2) for each cell was measured by Zen 2.3 lite software. 9 cells from non-epithelial tissue without non-specific signals were picked and measured as negative control. The intensity value of p63 channel per unit area for each epithelial cell was calculated, and adjusted background by subtracting the averaged p63 intensity value per unit area of negative controls. The normalized values of p63 intensity per unit area lower than 3.6 were scored as p63 negative; values between 3.6 to 20.0 were scored as p63 low; values higher than 20.0 were scored as p63 high. The percentiles of cells with different p63 fluorescent intensity were shown in the graph.

2-7. Quantification of Epithelial Cell Number per Unit Basement Membrane Length

20 lines along the basement membrane on cartilage side from each E18.5 WT and KO samples were drawn with their length recorded in Zen 2.3 lite software. The numbers of epithelial cells above the basement membrane lines were quantified. The bar graph showing the averaged number of epithelial cells per unit basement membrane length (mm). N=3 for both WT and KO. In total, cells along 7.13mm basement membrane were counted in WT; cells along 6.91mm basement membrane were counted in KO.

2-8. Foregut Culture

Foreguts were dissected from E8.5 *p63-CreERT2; R26-tdTomato* embryos (8 to 13 somites), exposed to 4-OHT (Sigma-Aldrich, H7904) for 2hrs and cultured on 6-well transwells (Costar 3450-Clear) in serum-free differentiation medium[111] supplemented with 0.02% Ascorbic Acid at 37C with 5% CO2. Cultures were monitored in a Zeiss Live Imaging System for 5 days and subsequently fixed (4% paraformaldehyde) for IF analysis.

2-9. H1N1 (PR8) Viral Preparation and Infection

The influenza-A (H1N1) mouse-adapted PR8 viral stock was prepared and amplified in 10-day-old embryonated chicken eggs. Allantoic fluid containing the virus was collected two days after egg inoculation, cleared from cellular debris by centrifugation, aliquoted and stored at -80°C for further use. The viral titer was determined as $1.2x10^9$ pfu/ml by plaque assay on Madin Darbin Canine Kidney cells using tenfold dilutions of allantoic fluid containing the virus. LD50 was initially determined in Balb/c mice as 450pfu. Different doses were tested in pilot experiments using WT C57Bl/6 mice, and 120pfu was found to achieve maximal generation of $Krt5^+$ pods with minimal mortality rates in both WT and *p63-CreERT2* mice.

Adult mice (2-3 months old) were anesthetized by isofluorane, and then intranasally administered with 120pfu of H1N1 virus diluted in 30ml PBS. Mock-infected animals received 30ml PBS. Successful infection was verified by weight loss and microscopy of lung frozen sections (massive infiltration of neutrophils and severe destruction of alveolar and airway structures).

2-10. Epithelial Cell isolation and FACS analysis

Tracheas and esophagus were isolated from adult mice with brief cleanup of stromal tissues. To isolate tracheal epithelial cells, only the main tracheal region below cricoid cartilage and above bifurcation was used to avoid confounding $p63^+$ cells from larynx or mainstem bronchi. Tracheal tubes were cut open and digested in Dispase Digestion Solution (16U/ml Dispase (Corning, 354235) + 10μg/ml DNase I diluted in PBS) at room temperature for 40min. Digestion was stopped by transferring tracheas to PBS. Epithelium was physically peeled off with forceps, and collected in Falcon tubes. Further digestion into single cells was done in Trypsin Digestion Solution (0.1% Trypsin + 3mM EDTA diluted in PBS) at 37°C for 10min. To isolate esophageal epithelial cells, the muscle layer surrounding esophageal epithelium was physically peeled off, following with Trypsin digestion of epithelium at 37°C for 10min. Digestion was stopped by addition of FACS buffer (HBSS + 2% BSA + 1x GlutaMax) with 10% FBS, followed by gentle pipetting and passage through 70μm cell strainer.

To stain for cell sorting, cells were suspended in staining buffer (FACS buffer + 10μg/ml DNase I), and incubated with antibodies for 40min at 4°C. Cells were washed with FACS buffer and DAPI was added to a final concentration of 1.25μg/ml before sorting. Sorting was performed on Influx (BD Biosciences) and

data analyzed with Flowjo (version 10). Cells were collected in staining buffer. The following antibodies were used: CD104-BV510 (1:50, BD Biosciences, 743079), EpCAM-APC (1:100, eBioscience, 17-5791-82), Lineage cocktail-Alexa Fluor 700 (1:20, Biolegend, 133313).

2-11. Bulk RNA sequencing and data analysis

BCs were isolated from adult mouse trachea and esophagus, E18.5 tracheal preBCs were isolated from embryonic tracheas, and these cells were sorted by DAPI$^-$ Lin$^-$ EpCAM$^+$ CD104hi tdTomato$^+$. For each replicate, three to five adult mice were pooled together for cell collection. For E18.5 tracheal preBCs, about 6 to 10 embryonic tracheas from 3 to 5 litters were pooled together for each replicate. 6 replicates for each cell population were sorted individually and prepared for sequencing. RNA was extracted using Direct-zol RNA MicroPrep kit (Zymo Research, R2060). RNA integrity was confirmed by Bioanalyzer. Samples with RNA integrity numbers >9.0 were sent for sequencing.

The sample numbers were: adult tracheal BCs (CY001, CY004, CY005, CY010, CY011, CY012); adult esophageal BCs (CY002, CY003, CY006, CY013, CY014, CY015); E18.5 tracheal preBCs (CY007, CY008, CY009, CY016, CY017, CY018). Library preparation and RNA sequencing were performed by JP Sulzberger Columbia Genome Center. mRNAs were enriched using poly-A pull-down from total RNA samples. Library preparation was conducted using Illumina TruSeq RNA prep kit.

For CY001-CY006, libraries were sequenced using Illumina HiSeq4000 at Columbia Genome Center, which yielded target number of single-end 100bp 25-30 million reads for each sample. RTA (Illumina) was used for base calling and bcl12fastq2 (version 2.17) was used for converting BCL to fastq format, coupled with adaptor trimming.

For CY007-CY018, libraries were sequenced using Illumina NovaSeq 6000 at Columbia Genome Center, which yielded target number of paired-end 100bp 20 million reads for each sample. RTA (Illumina) was used for base calling and bcl12fastq2 (version 2.20) was used for converting BCL to fastq format, coupled with adaptor trimming.

For all samples (CY001-CY018), a pseudoalignment was performed to a kallisto index created from mouse transcriptome GRCm38 using kallisto (0.44.0). Differential gene expression analysis was performed using the R package Sleuth, and visualized in volcano plots. Specifically, Eso- and Tr-specific over-

expressing genes, as determined and highlighted in the volcano plots, were used as GSEA gene sets to determine Eso- and Tr-scores in single cells, as described in the following section.

2-12. Single cell RNA sequencing and data analysis

Cells were sorted into staining buffer and stored on ice until proceeding to the 10X Genomics single cell platform at Columbia Genome Center. Single Cell 3' libraries were prepared using the Chromium Single Cell 3' v2 Protocol (CG00052) according to the manufacturer's manual (10X Genomics). The pooled, 3'-end libraries were sequenced using Illumina HiSeq4000. Cell Ranger version 2.1.1 was used for primary data analysis, including demultiplexing, alignment, mapping, gene expression quantification, dimension reduction analysis and clustering analysis within individual datasets. Specifically, for alignment and mapping, the mm10 reference genome and corresponding annotation were used.

For secondary protein activity analysis, single cell gene regulatory networks, including esoBC- and trBC-specific ones, were constructed using the ARACNe algorithm[112]. ARACNe was run with 100 bootstrap iterations using 1813 transcription factors and 969 transcriptional cofactors. Parameters were set to zero DPI (Data Processing Inequality) tolerance and MI (Mutual Information) p-value (using MI computed by permuting the original dataset as null model) threshold of 10^{-8}. Protein activity profiles were predicted using the VIPER algorithm[113] based on individual ARACNe network, or using the metaVIPER algorithm[114] based on both esoBC and trBC ARACNe networks. For protein activity-based clustering analysis, iterClust[115], iterative clustering algorithm, was used.

For tertiary cross-datasets analysis, single cell gene expression matrices generated from primary analysis were combined, followed by tSNE (t-Distributed Stochastic Neighbor Embedding) dimension reduction analysis and DBSCAN (Density-based Spatial Clustering of Applications with Noise) clustering analysis. Eso- and Tr-scores were quantified using GSEA analysis, with Eso- and Tr-specific over-expressing genes as gene sets. Such over-expressing genes were determined from bulk RNA-Seq analysis, as described in the previous section.

2-13. Quantification and Statistical Analysis

Quantification and statistical analysis have already been detailed in the Methods section above, associated with each experiment, as well as in the figure legends. All quantification for colocalization and marker analyses were performed in Adobe Photoshop CS6. Statistical analyses were performed in Microsoft Excel or GraphPad Prism 7. As shown in figure legends, data in graphs were shown in mean ± SEM. Statistical significance of differences in three or more groups of data were determined by one-way ANOVA and post hoc Tukey's test. Statistical significance of differences between two groups of data were determined by unpaired two-tailed t test.

Chapter 3. Two Sequential Lineage Segregation Events of p63^{+} Progenitor Cells During Airway Development

3-1. p63 expression pattern in mouse respiratory system development

p63 gene shows distinct expression pattern in tracheal epithelium during airway development. Before E14.5, p63 is broadly and weakly expressed in majority of tracheal epithelial cells (~70%). These cells are considered as immature p63^{+} Krt5^{-} cells as they do not have characteristic features of mature BCs (Figures 1A-B; Figure 2B). Moreover, they are identified in both luminal and basal locations, so they are different from the ITGB4^{hi} prebasal cells described before[9]. The p63 signal intensity vary greatly among these cells, suggesting a remarkable heterogeneity in the early p63^{+} progenitor population (Figure 1C). E15.5 represents a critical transition time point as some p63^{+} cells begin to express Krt5 at low levels and gradually change the nuclear polarity, being described as nascent preBCs. This is the first time that morphological segregation between basal versus luminal compartments is evident, and about 60% of total epithelial cells are positive for p63 signal (Figures 1A-B). From E16.5 to E18.5, relatively high level of p63 is co-expressed with Krt5 in the cells (~24% of total epithelial cells) which assume basal location with horizontal nuclear orientation (Figures 1A-B). These cells were referred as preBCs to distinguish from their postnatal counterparts, as they still have to acquire addition BC markers such as NGFR in postnatal growth.

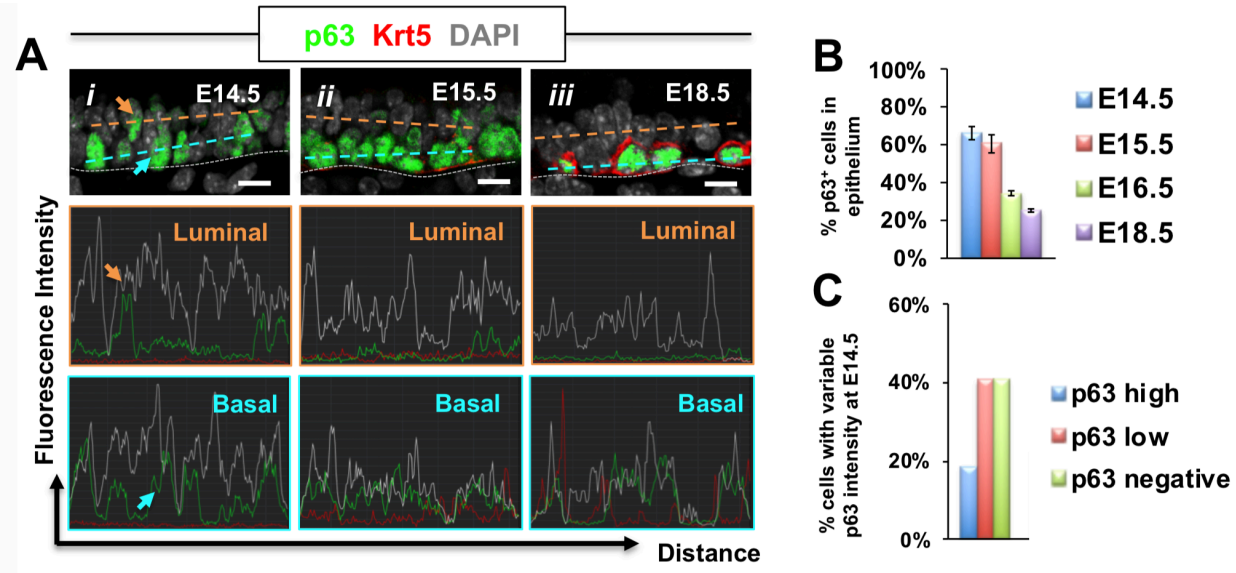


Figure 1. Dynamics of p63 expression pattern in tracheal development.

(A) Segregation of basal and luminal compartments in the pseudostratified trachea, expression pattern analysis of p63 (green peaks) and Krt5 (red peaks) in both basal and luminal locations at E14.5, E15.5 and E18.5. Orange arrow showing luminal $p63^+$ cells with high intensity signals. Cyan arrow showing basal $p63^+$ cells with high intensity signals. **(B)** Graph showing % $p63^+$ cells in tracheal epithelium at indicated stages. **(C)** Graph showing % p63-expressing cells with variable fluorescence intensities at E14.5 from 150 randomly selected epithelial cells in trachea. Scale bars: 10μm.

To identify the onset of p63 expression in respiratory progenitors (marked by Nkx2-1), we searched for the earliest p63-expressing cells during initiation of trachea/lung development in *Nkx2-1-GFP* embryos in which the earlies $Nkx2\text{-}1^+$ cells could be detected using GFP expression. Immunofluorescence (IF) first detected a lot of $p63^+$ cells intercalated within the GFP^+ respiratory domain at E9.0 before the evagination of primary lung buds (Figure 2A). At E9.5, when the two primary lung buds evaginated from the ventral foregut, a small population of $p63^+$ GFP^+ cells were found in the tracheal primordium and the scattered proximal regions of the early lung buds (Figure 2A). A day later $p63^+$ GFP^+ cells were mostly confined to the tracheal region, where it remained abundant in subsequent stages (Figures 2A-B). Therefore, the onset of p63 expression in lung development was determined not from E10.5, but instead from E9.0[10], and a novel p63-expressing population inside the primary lung buds was revealed during the specification of respiratory domain.

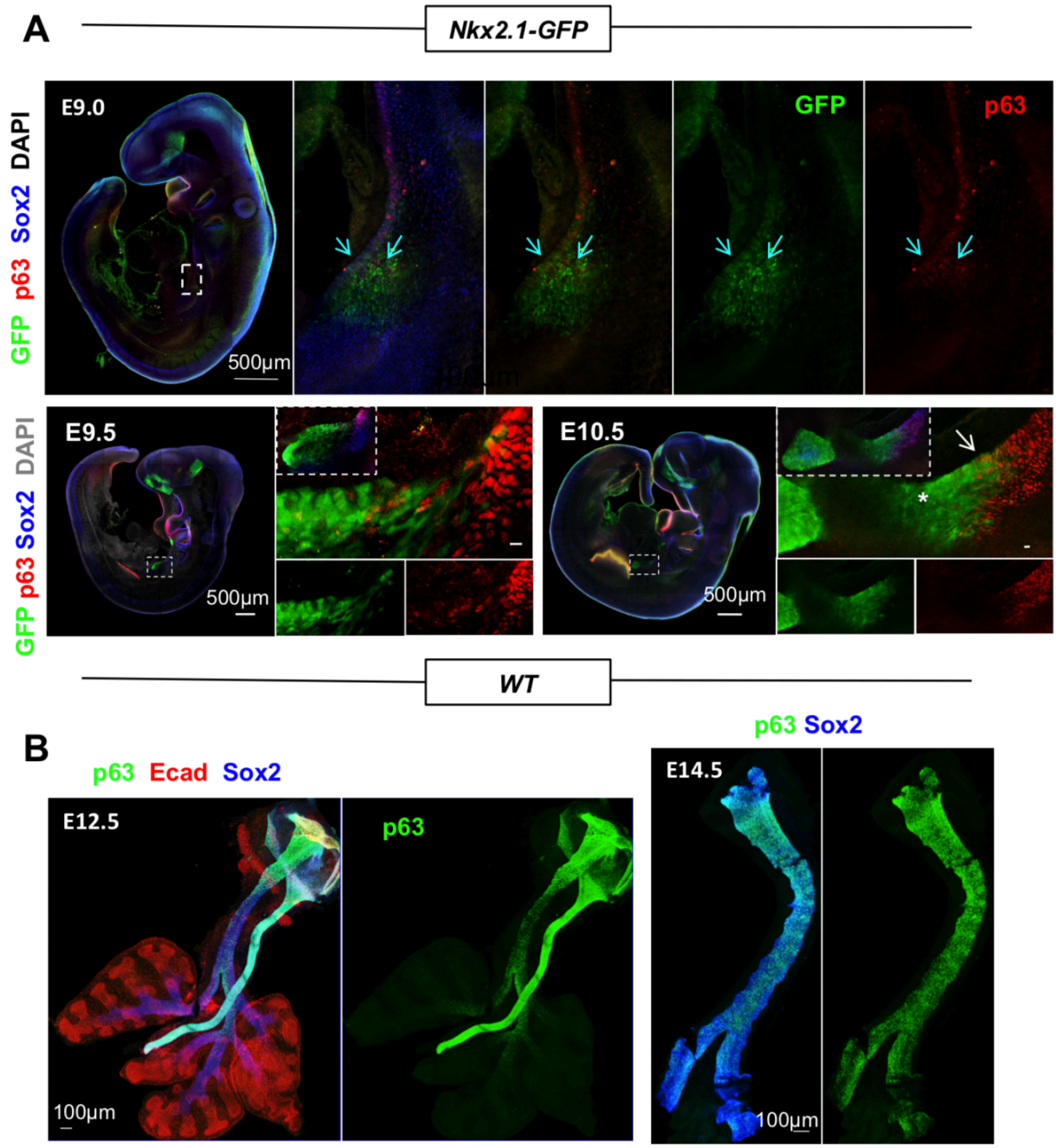


Figure 2. Unexpected early p63 expression in the respiratory domain.

(A) Immunofluorescence (IF) of E9.0, E9.5 and E10.5 *Nkx2.1-GFP* embryos; boxed areas enlarged in right panels. GFP^+ $p63^+$ cells in lung (cyan arrows) or tracheal (white arrow) primordia. (*) GFP^+ $p63^-$ in lung. **(B)** Whole-mount IF of E12.5 trachea & lung (left) and E14.5 trachea (right). Scale bars: 10μm unless noted.

3-2. preBCs at E18.5 give rise to adult BC pool in trachea and generate SMGs

preBCs at E18.5 in tracheal epithelium are characterized by a well-defined basal layer of strongly labeled $p63^+$ $Krt5^+$ cells (Figure 3A). To determine whether preBCs have similar potentials as their postnatal counterparts, we performed lineage tracing using *p63-CreERT2; R26-tdTomato* mouse line with TM administration at E17.5 by oral gavage, and followed the fate of lineage-labeled preBCs at different postnatal stages (Figure 3A). Exposure of *p63-CreERT2* embryos to TM at E17.5 labeled 96% of the $p63^+$ $Krt8^-$ preBCs at E18.5 and labeling frequency was maintained high throughout adulthood (postnatal days P7, P21) (Figures 3B & C). Labeled luminal cells ($p63^-$ $Krt8^+$ $tdTom^+$) increased significantly to 23% of the total tracheal epithelium at P21, confirming that the E18.5 lineage-labeled prebasal cells self-renew and simultaneously generate luminal descendants, resembling their postnatal BC counterpart (Figures 3B & C). Therefore, preBCs are direct precursors of adult BCs and they have higher turnover rate to accommodate relatively faster postnatal growth.

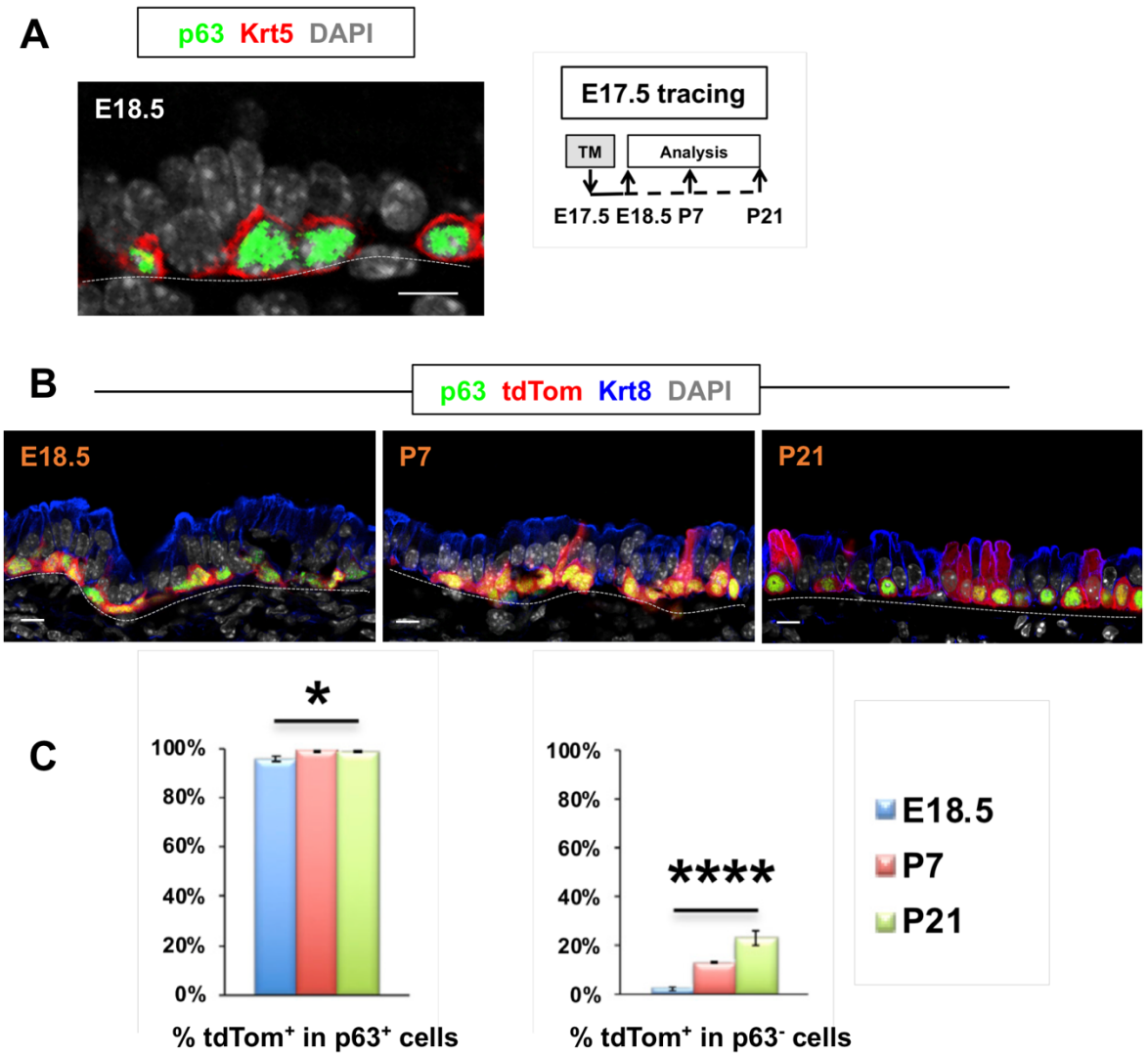


Figure 3. E18.5 preBCs generated adult BC pool and simultaneously contributed to luminal compartment during postnatal growth.

(A) Labeling of E18.5 preBCs in *p63-CreERT2; R26-tdTomato* mice with TM exposure at E17.5. **(B)** IF of tracheal sections showing tdTom labeling in $p63^+$ $Krt8^-$ basal compartment and $p63^-$ $Krt8^+$ luminal compartment at E18.5, P7 and P21. **(C)** Left graphs: % $tdTom^+$ $p63^+$ cells in total $p63^+$ population; right graphs: % $tdTom^+$ $p63^-$ cells in total $p63^-$ population at each stage. Graphs: mean ± SEM from 6-12 fields per sample, N⩾3 per stage. Statistics: one-way ANOVA, *P<0.05; ****P<0.0001. Scale bars: 10μm.

SMGs are small epithelial appendages embedded in the interstitial layer of the SAE. Although in mice, SMGs are restricted to the most proximal part of trachea, in humans they were found throughout the cartilaginous airways deeply inside the lung. SMGs are derived from progenitors reside in the SAE during postnatal development[116]. Initially progenitor cells in $Sox2^+$ PGPs (primordial glandular placodes) receive stimulatory signals such as WNT/β-catenin, evaginate into the stroma and further branch into the network of SMG[117]. However, which specific cell types in the SAE give rise to SMG in development remains unknown. To explore the ontogeny of SMGs, airway preBCs at E18.5 were hypothesized as the precursors for SMGs. The lineage tracing strategy was described above with TM exposure at E17.5 (Figure 4A). preBCs in SAE at E18.5 were specifically and efficiently labeled. At E18.5, tdTomato already labeled $p63^+$ PGP-like structures in the upper trachea (Figure 4B). At P7 and P21, extensive labeling was observed in the nascent airway SMGs (which are double positive for both Nkx2-1 and Sox2), supporting the hypothesis that SMGs are derived from $p63^+$ preBCs in SAE at E18.5 (Figure 4C & D). This developmental relationship between SMGs and SAE BCs may underlie the remarkable cellular plasticity of SMG myoepithelial cells in regenerating damaged SAE in adulthood[96,97].

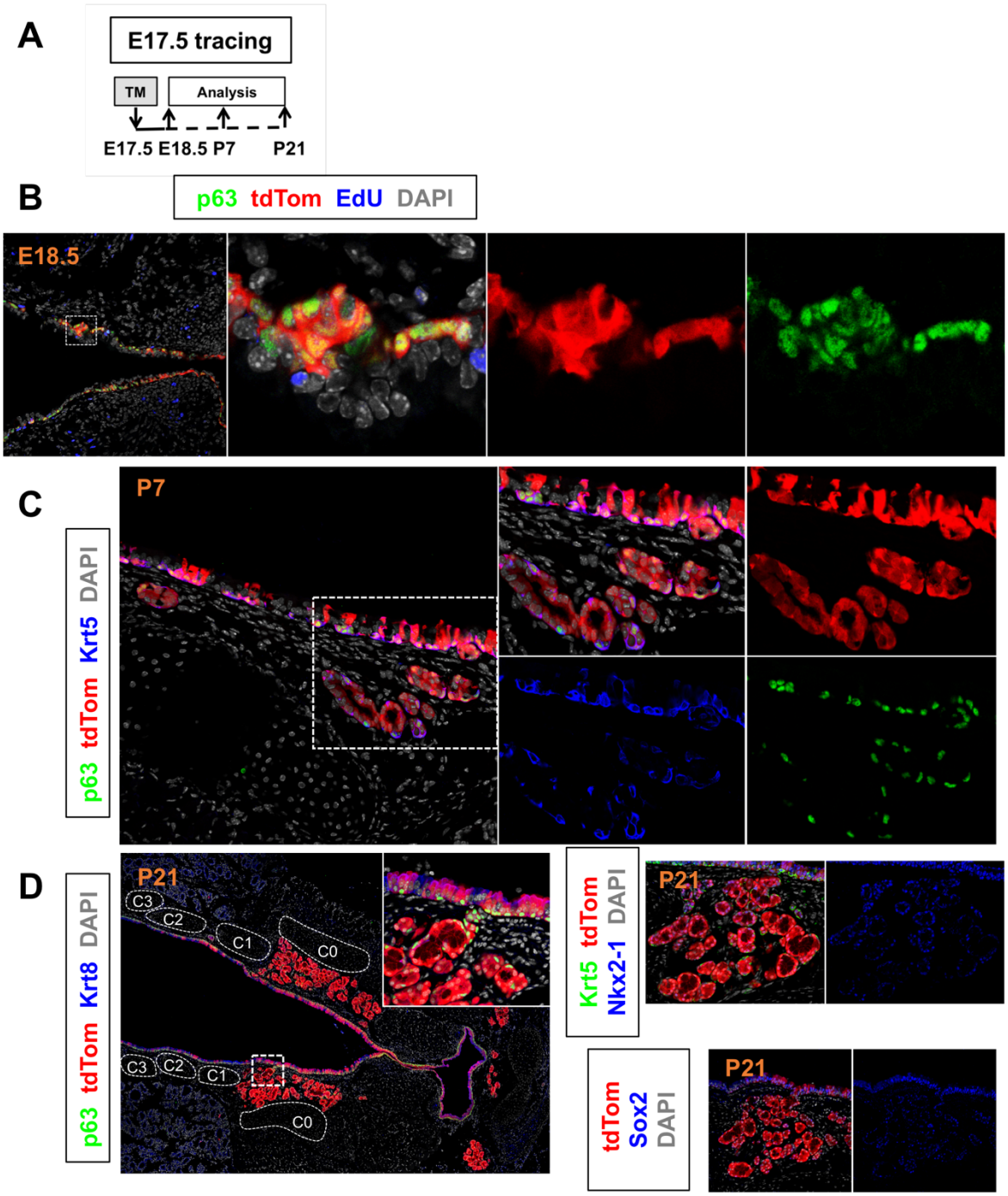


Figure 4. E18.5 preBCs in trachea generated SMG during postnatal growth.

(A) Lineage tracing of *p63-CreERT2; R26-tdTomato* embryos with TM exposure at E17.5. **(B-D)** IF of top tracheal sections showing tdTom labeling in developing SMGs.

3-3. Tracheal BC pool is established at E13.5 by a subpopulation of p63$^+$ cells

To investigate the ontogeny of tracheal BCs, lineage tracing of p63$^+$ progenitors was performed at distinct embryonic stages, and the lineage-labeling frequencies for E18.5 preBCs were evaluated, which further gave rise to adult BC pool. In spite of differences in recombination from E17.5 tracing, TM at E13.5 or E14.5 both resulted in labeling of >84% of the E18.5 p63$^+$ Krt8$^-$ prebasal cells. This could be ascribed to the strong p63 promoter activity in some of these cells (Figures 5 & 6). The data suggested that the pool of multipotent precursors destined to become BCs was largely established around E13.5, even preceding the appearance of Krt5, one of the earliest markers of initiation of BC program (Figures 5A & 6A). Consistent with this, TM exposure before E13.5 resulted in markedly reduced number of p63$^+$ tdTom$^+$ cells at E18.5 (ex. only 57.9% ± 10.6% with TM treatment at E12.5). Interestingly, the labeling frequency in luminal compartment analyzed at E18.5 increased significantly when mice were exposed to TM at earlier stages (TM at E14.5: 18.1% ± 1.9%; TM at E13.5: 33.7% ± 3.2%), suggesting that at least some of the p63$^+$ tdTom$^+$ cells could have been committed to luminal differentiation before the basal compartment was established (Figures 5C & 6C).

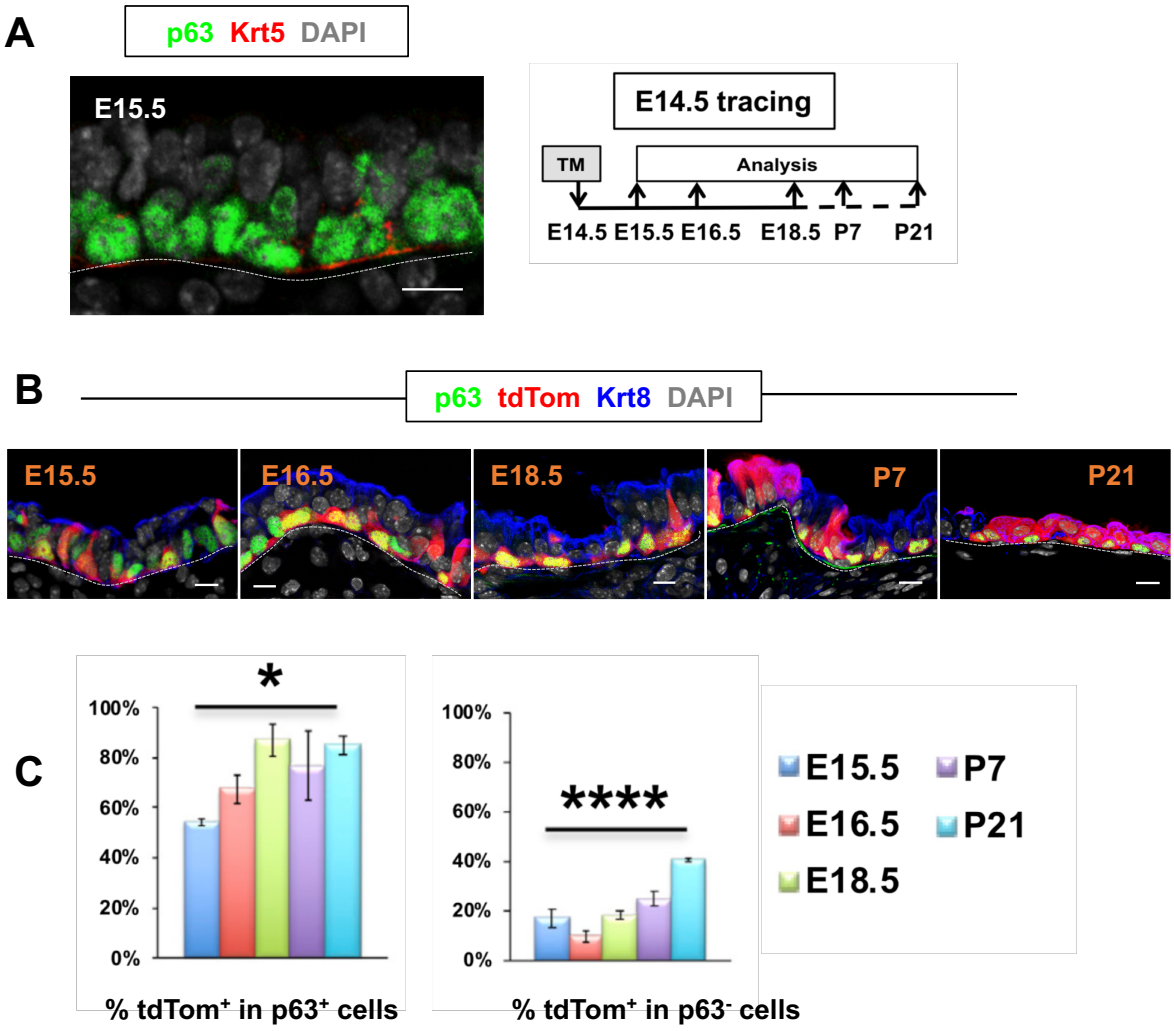


Figure 5. E15.5 nascent preBCs generated adult BC pool and contributed significantly to luminal compartment during postnatal growth.

(A) Labeling of E15.5 preBCs in *p63-CreERT2; R26-tdTomato* mice with TM exposure at E14.5. **(B)** IF of tracheal sections showing tdTom labeling in $p63^+$ $Krt8^-$ basal compartment and $p63^-$ $Krt8^+$ luminal compartment at designated stages. **(C)** Left graphs: % $tdTom^+$ $p63^+$ cells in total $p63^+$ population; right graphs: % $tdTom^+$ $p63^-$ cells in total $p63^-$ population at each stage. Graphs: mean ± SEM from 6-12 fields per sample, N⩾3 per stage, except P7 (N=2). Statistics: one-way ANOVA, *P<0.05; ****P<0.0001. Scale bars: 10μm.

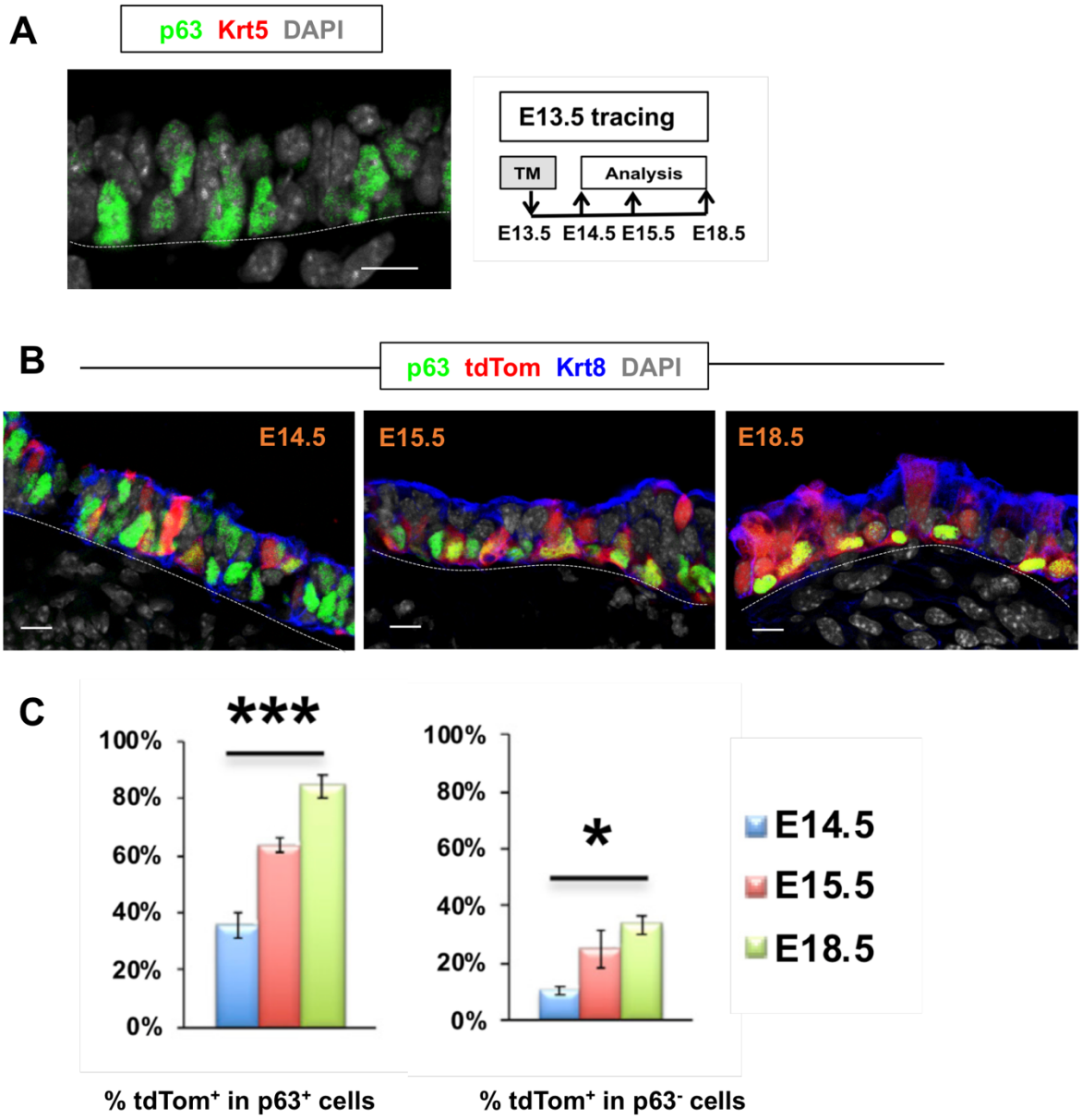


Figure 6. Immature p63$^+$ Krt5$^-$ progenitors labeled at E13.5 gave rise to majority of preBCs at E18.5 and contributed significantly to luminal compartment in the meanwhile.

(A) Labeling of E14.5 immature p63$^+$ Krt5$^-$ progenitors in *p63-CreERT2; R26-tdTomato* mice with TM exposure at E13.5. **(B)** IF of tracheal sections showing tdTom labeling in p63$^+$ Krt8$^-$ basal compartment and p63$^-$ Krt8$^+$ luminal compartment at designated stages. **(C)** Left graphs: % tdTom$^+$ p63$^+$ cells in total p63$^+$ population; right graphs: % tdTom$^+$ p63$^-$ cells in total p63$^-$ population at each stage. Graphs: mean ± SEM

from 6-12 fields per sample, N⩾3 per stage. Statistics: one-way ANOVA, *P<0.05; ***P<0.001. Scale bars: 10μm.

Next, to investigate whether the population of $p63^+$ $tdTom^+$ cells expanded preferentially compared to the other cell types, EdU was administered simultaneously with TM to *p63-CreERT2* embryos at E13.5, E14.5, and E17.5 and incorporation was quantified 24hr later in $p63^+$ $tdTom^+$ or $tdTom^-$, and $p63^-$ cells. As expected, EdU incorporation declined with age in all groups (Figure 7A). Interestingly, no evidence of preferential EdU incorporation was found in any of the $p63^+$ populations (Figure 7A). Therefore, p63 expression does not confer proliferative advantage at least in the developing trachea. The data also suggest that the above lineage labeling was not biased towards a higher proliferative $p63^+$ subpopulation.

This lineage tracing data suggested that BC precursors were preferentially labeled over luminal progenitors with TM administration at E13.5. There are two potential explanations for this result. One is that newly specified BCs continue to undergo Cre-mediated recombination due to a TM residual effect. The second explanation is that the recombination events predominantly occurred in BC precursors, due to their high p63 promoter activity. To distinguish between these possibilities, TM dose was reduced (from 160 μg/g body weight to 10 μg/g and 4 μg/g) to minimize residual effect; this resulted in major decrease in efficiency of labeling at E14.5 from 35.7% to 6.3% and 0.6%, respectively (Figures 7B). However, analysis of E18.5 tracheas showed that in spite of the reduced recombination efficiencies at lower doses, the relative percentiles of $tdTom^+$ $p63^+$ BCs in the total lineage-labeled population were consistently increased (Figure 7B). These suggested that BC precursors are preferentially labeled in this lineage tracing system, but not due to TM residual effect. Thus, the airway BC pool is established by a group of $p63^+$ progenitors presumably with relatively higher p63 promoter activity at E13.5, earlier than the morphological segregation of basal/luminal compartments and Krt5 expression from E15.5.

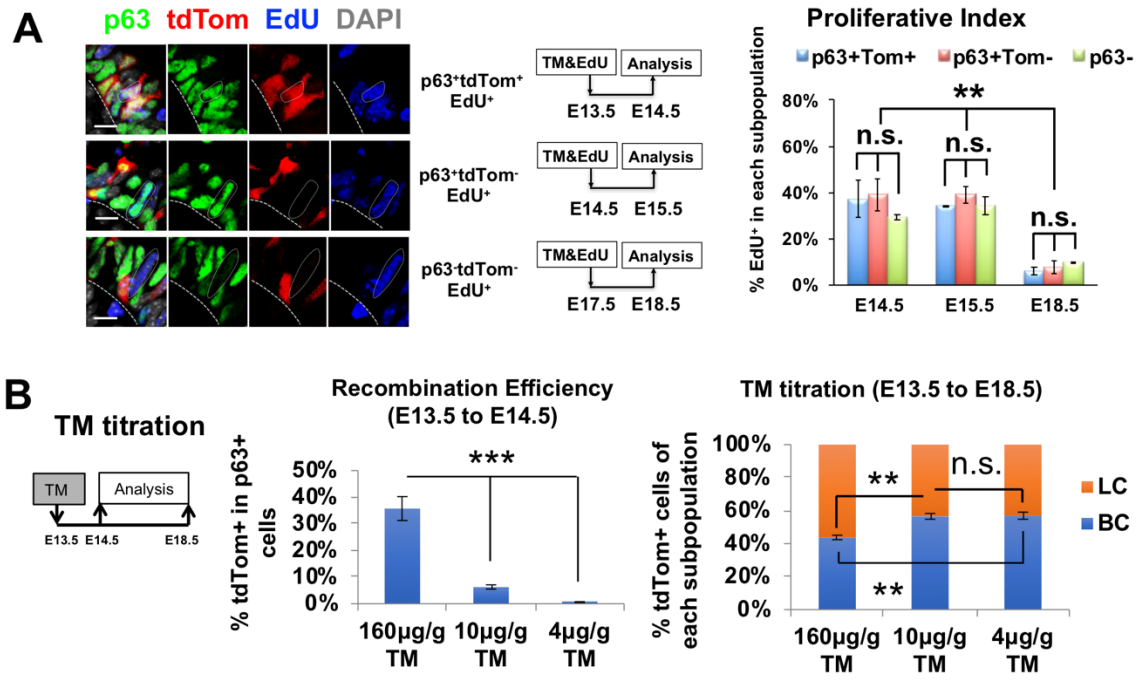


Figure 7. E13.5 tracing preferentially labeled BC precursors, not due to labeling bias of proliferative activity or TM residual effect.

(A) *p63-CreERT2; R26-tdTomato* mice exposed to EdU and TM simultaneously at E13.5, E14.5 or E17.5. EdU incorporation and lineage-labeling analysis in tracheal sections after 24hr: $p63^+$ ($tdTom^+$ or $tdTom^-$) or $p63^-$ cells (representative images and quantification). **(B)** TM titration experiment showed dramatic decreased recombination efficiency with decreased TM dosage. However, the labeling frequency of basal compartment increased significantly with lower TM dosage, excluding the possibility of TM residual effect. Statistics: one-way ANOVA, **P<0.01; ***P<0.001; n.s. non-significant. Scale bars: 10μm.

3-4. Embryonic $p63^+$ progenitors contribute equally to secretory and multiciliated lineages during tracheal development

To further investigate the contribution from p63 lineage to multiciliated and secretory lineages, E13.5 tracing was analyzed during embryonic development, as the confounding conversion from secretory to multiciliated lineage is limited. When pregnant mice were exposed to tamoxifen at E13.5, 34% of total luminal cells carried the lineage-label at E18.5 (Figure 6C). Interestingly, quantification for the lineage-labeled $Foxj1^+$ multiciliated cells and $CC10^+$ secretory cells showed similar proportions (37% for multiciliated lineage while

36% for secretory lineage) (Figure 8A). This indicated that lineage-labeled embryonic p63^{+} cells underwent a balanced differentiation program, as they made comparable contributions to multiciliated and secretory lineages during development. Furthermore, E14.5 tracing showed a significant increase in the lineage-labeled A-tub^{+} or CC10^{+} luminal cells at late prenatal and postnatal stages (Figure 8B). Thus, embryonic p63^{+} progenitors make equal and considerable contributions to luminal secretory and multiciliated lineages during tracheal development.

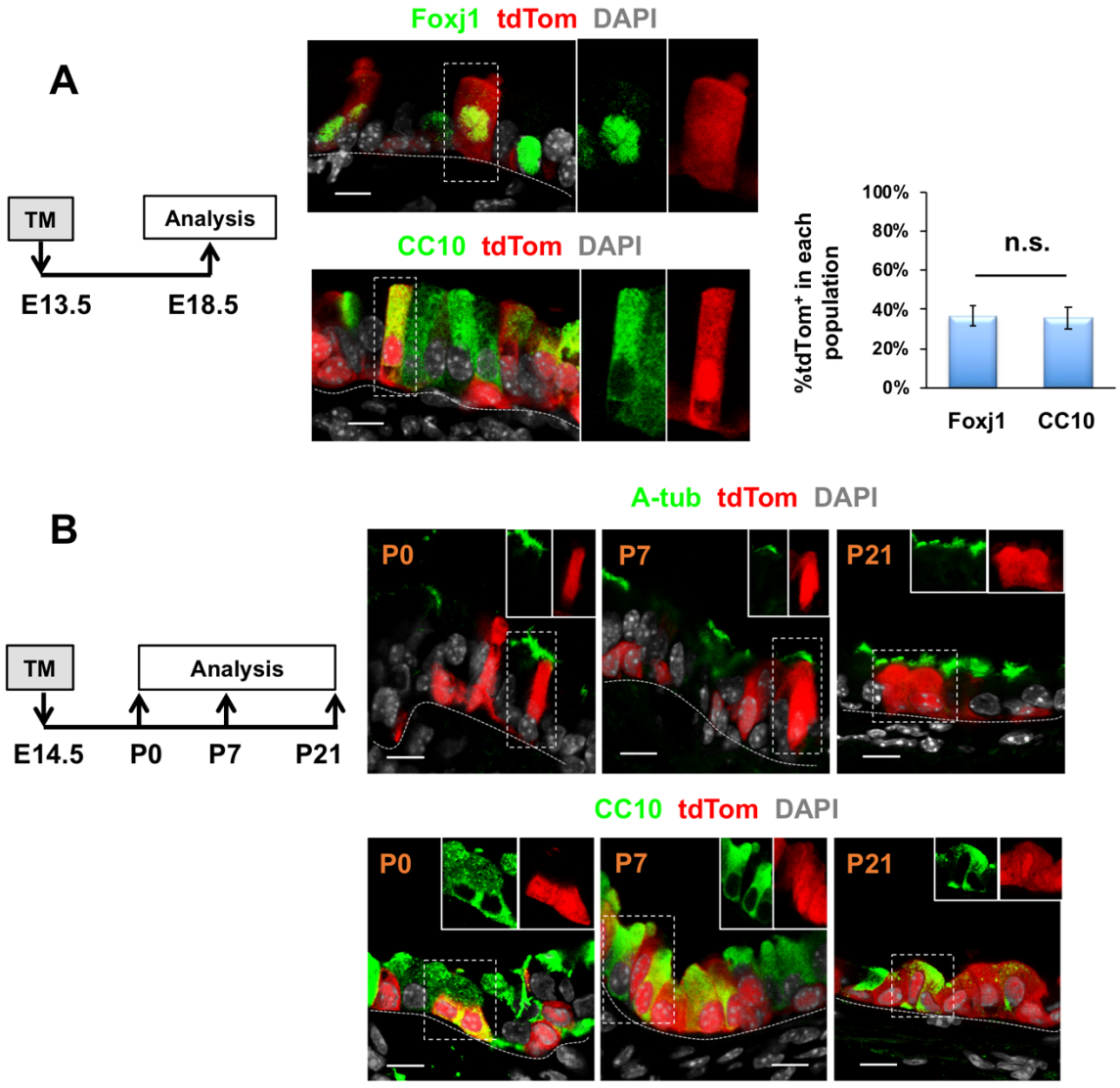


Figure 8. Lineage restriction of p63^{+} cells contributed equally to multiciliated and secretory lineages in developing trachea.

(A-B) Immunofluorescence (IF) of tracheal sections from *p63-CreERT2; R26-tdTomato* mice. Contribution of embryonic p63 to luminal tracheal secretory and multiciliated cells pre- and postnatally. TM exposure at E13.5 or E14.5 and analysis at E18.5, P0-P21. tdTom double-labeled with differentiation marker; dotted area shown as single channel in lateral panels or insets. Graph: % $tdTom^+$ cells in multiciliated and secretory cells at E18.5. Graph: mean ± SEM from 6-12 fields per sample, N⩾3 per stage. Statistics: Student's t test, n.s. non-significant. Scale bars: 10μm.

3-5. $p63^+$ progenitors at the onset of lung specification are multipotent to generate both airway and alveolar lineages

As described above, p63-expressing progenitors are already present in the primitive respiratory domain at the onset of lung specification since E9.0. To lineage-trace p63 at the earliest stages observed, E8.5, E9.5 or E10.5 embryos were exposed to TM (160 μg/g, maternal oral gavage). Analysis of E18.5 tracheas showed extensive tdTom labeling in the pseudostratified epithelium at these stages, confirming the contribution of these progenitors from as early as E8.5 (Figures 9A & B; 10A-D). Intrapulmonary airways were nearly unlabeled in E10.5 TM-treated embryos (Figure 9B). Surprisingly, TM exposure at E9.5 or E8.5 resulted in abundant $tdTom^+$ cells in E18.5 intrapulmonary airways extending ectopically to alveolar saccules (Figures 9A & 10D). IF showed $tdTom^+$ cells double-labeled with markers of airways (CC10: secretory; Cgrp: neuroendocrine) and alveolar (Pdpn: type I; pro-Spc: type II) cells (Figures 9A & 10D). E8.5 TM-treated embryos showed the broader distribution of $tdTom^+$ cells extending to the lung domain at different stages from E9.5-E18.5 (Figures 9A-D). To rule out putative TM residual effect *in vivo*, E8.5 foregut explants from *p63-CreERT2; R26-tdTomato* embryos were isolated, treated with 4-hydroxytamoxifen (4-OHT) for 2hrs and subsequently cultured for 5 days. IF showed the ectopic tdTom extending to the distal $Nkx2.1^+$ $Sox9^+$ domain of lung buds, confirming the pattern observed with E8.5 TM administration in vivo (Figure 9E). Thus, in the early respiratory tract, p63 marks multipotent progenitors able to generate potentially proximal and distal epithelial cell components, as predicted for other progenitors occupying the Nkx2.1 domain. This ability is lost by E10.5, when they become lineage-restricted to tracheal and proximal intrapulmonary airways.

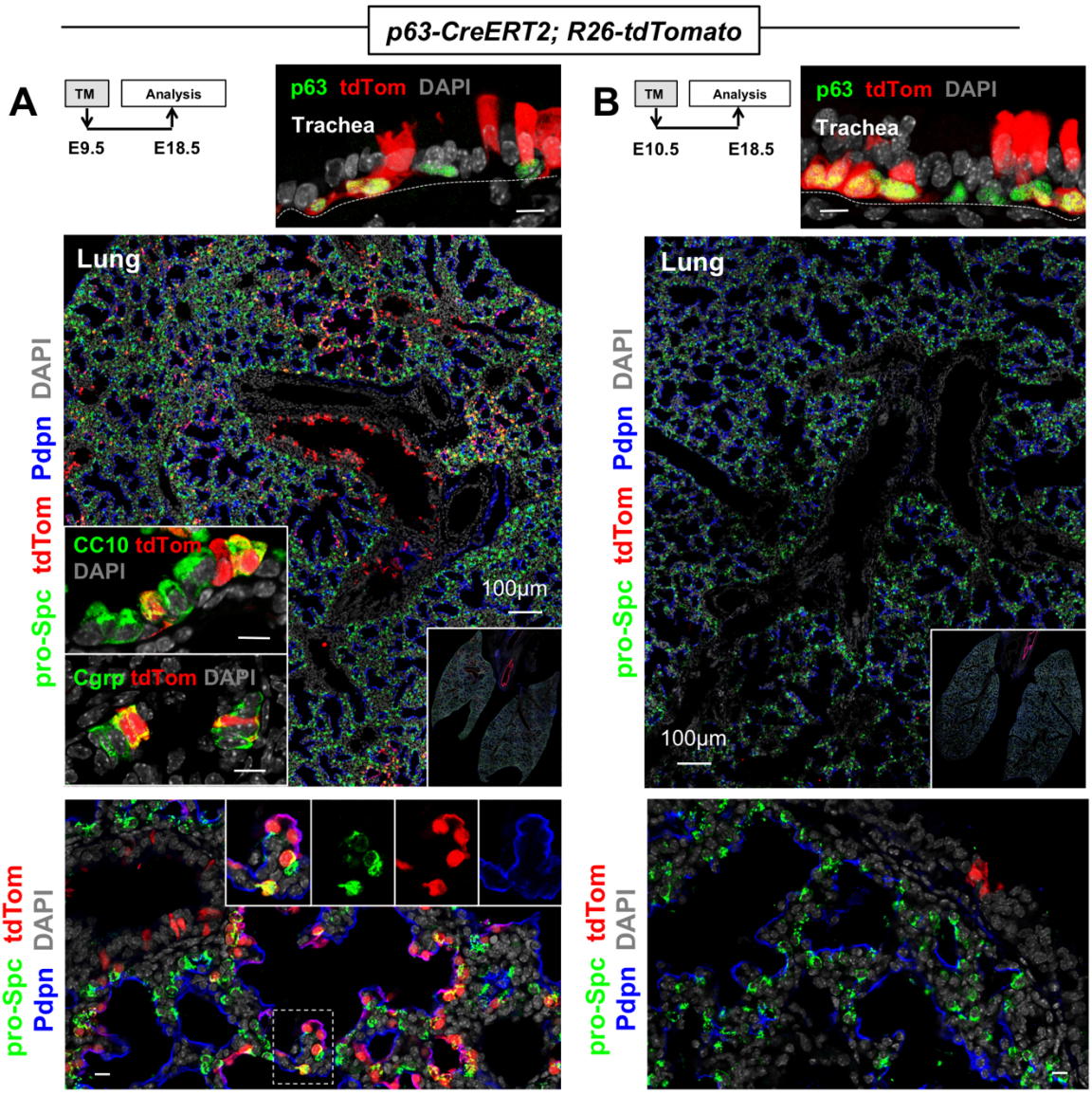


Figure 9. p63 labels multipotent progenitors of airways and alveoli and later become lineage-restricted to airways.

p63-CreERT2; R26-tdTomato lineage labeling. E18.5 trachea (top panels) and lungs (bottom panels: intrapulmonary airways and alveolar saccules) from embryos exposed to TM at E9.5 **(A)** or E10.5 **(B)**. tdTom^{+} double-labeling with cell type-specific differentiation marker: insets depict secretory (CC10), neuroendocrine (Cgrp), alveolar type 1 (Pdpn) and type 2 (pro-Spc) cells. Labeling pattern similar in trachea but dramatically different in lung indicative of lineage restriction. E9.5 tracing: N=11 embryos from 3 litters,

all with extensive tdTom labeling in alveolar compartment; E10.5 tracing: N=7 embryos from 2 litters, all with minimal tdTom labeling in alveolar compartment. For TM exposure after E10.5: N>18 from at least 8 litters, none with alveolar labeling. Scale bars: 10μm unless noted.

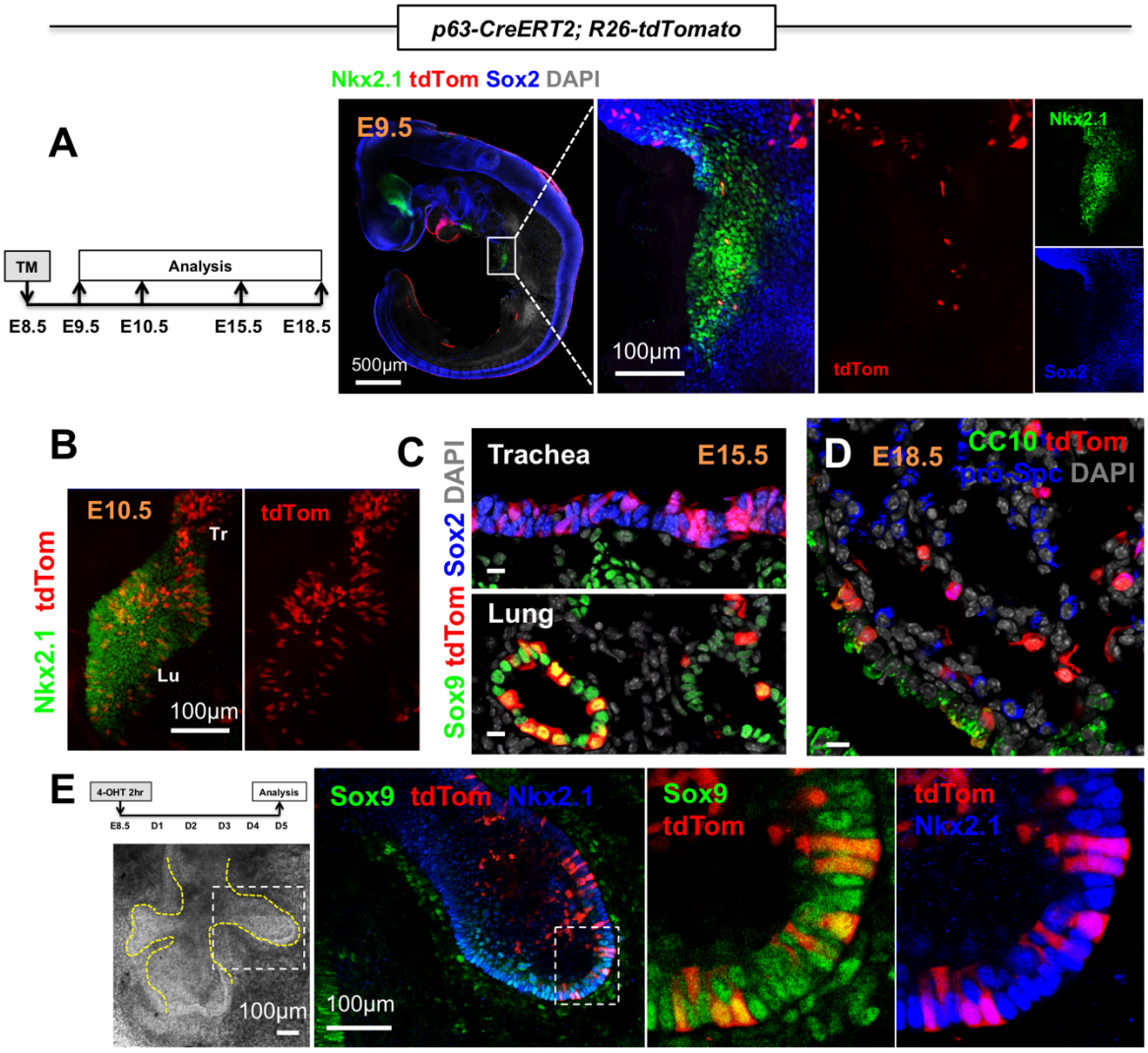


Figure 10. Lineage restriction of extrapulmonary p63⁺ multipotent progenitors in the lung.

(A-E) *p63-CreERT2; R26-tdTomato* lineage labeling: TM administration at E8.5 *in vivo* **(A-D)** or *ex-vivo* **(E)**. IF of E9.5-E18.5 lungs: Nkx2.1⁺ tdTom⁺ cells in tracheal (Tr) and lung (Lu) primordia; at E15.5 tdTom⁺ cells are abundant in tracheal epithelium but extending ectopically to distal (Sox9⁺) lung buds (**C**, lower panel), by E18.5 tdTom⁺ are present in intrapulmonary airways (double-labeled with CC10) and alveolar sacs (**D**, double-labeled with pro-Spc). N=6 embryos from 3 litters showing extensive alveolar labeling with TM

exposure at E8.5. **(E)** Foregut explants from E8.5 reporter treated with 4-OHT and cultured for 5 days (left). IF: $tdTom^+$ distal lung buds double labeled with Sox9 or Nkx2.1. Bud with ectopic distal tdTom labeling (box) enlarged in right panels. N=8 explants from two independent experiments. Scale bars: 10μm unless noted.

To learn whether intrapulmonary airways were lineage-restricted at comparable early stages, lineage tracing of Sox9, which, similar to Id2, Spc and others, marks distal lung epithelial buds at E10.5 was performed[40]. *Sox9-CreERT2; R26-tdTomato* embryos were exposed to TM between E9.5-E11.5. Analysis of TM-treated at E11.5 embryos showed expected distribution of tdTom selectively in intrapulmonary (bronchioles) but not in tracheal epithelium. By contrast, TM at E9.5 resulted in tdTom extending ectopically to the tracheal epithelium (Figure 11A). Intriguingly, when lineage labeling was induced at E9.5, all the $CGRP^+$ neuroendocrine cells in neuroendocrine bodies (NEBs) were $tdTom^+$ (Figure 11C). However, in E11.5 tracing, not all the neuroendocrine cells in one NEB were labeled consistently, suggesting a multi-clonal origin for NEBs (Figure 11B). Additionally, both distal progenitors and proximal progenitors may contribute to the neuroendocrine lineages inside the lung.

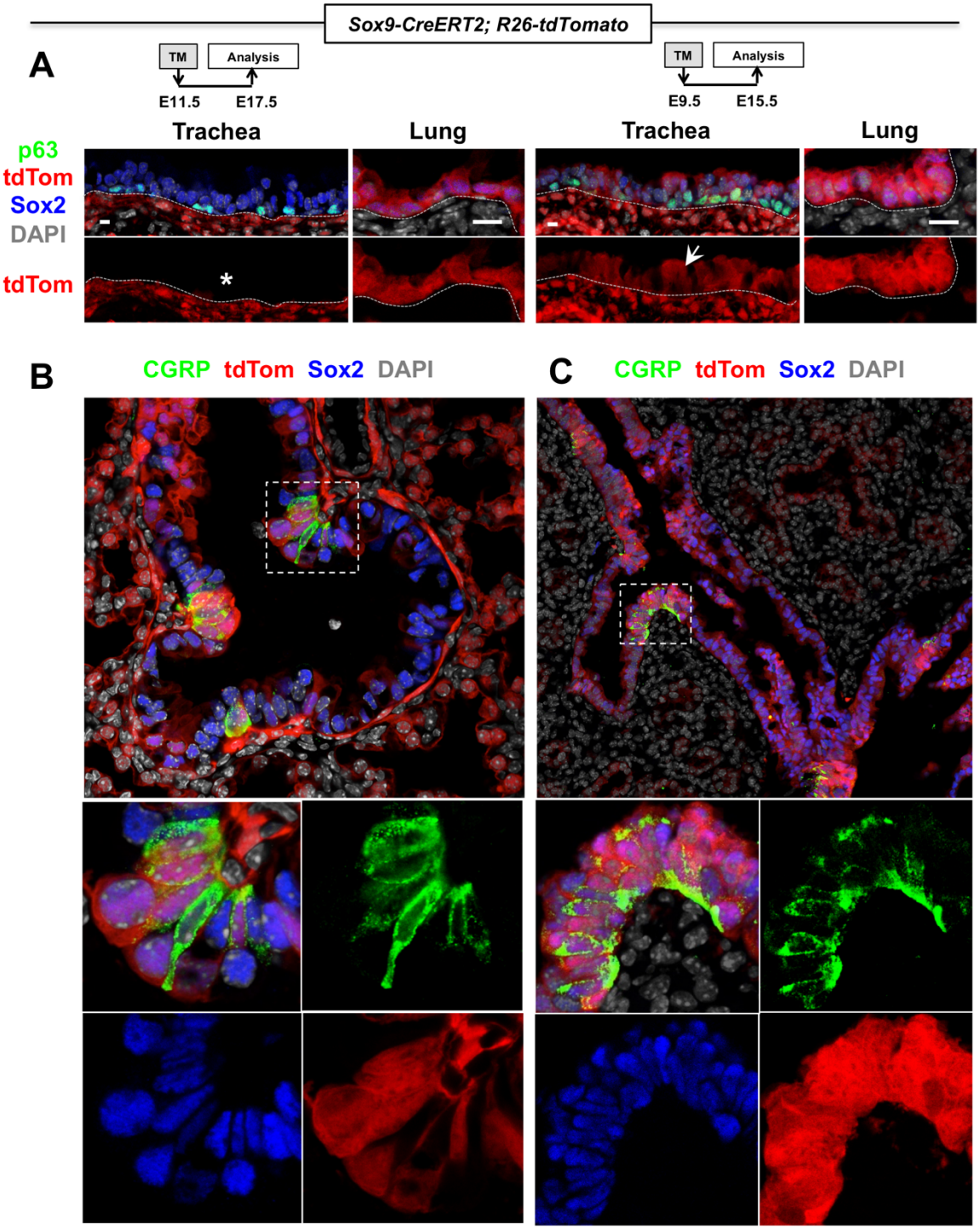


Figure 11. Lineage restriction of Sox9^{+} intrapulmonary multipotent progenitors in the lung.

(A-C) Lineage analysis of *Sox9-CreERT2; R26-tdTomato* mice *in vivo*. IF: E18.5 trachea and lung (intrapulmonary airways) from embryos exposed to TM at E11.5 **(A** left panel & **B)** or E9.5 **(A** right panel & **C)**. Proper lineage restriction of distal Sox9 descendants to intrapulmonary airways confirmed by absence

of tdTom$^+$ in tracheal epithelium (* p63$^+$Sox2$^+$) of TM-treated embryos at E11.5 but not at E9.5 (arrow: tdTom labeling extending to trachea). **(B)** Not all the neuroendocrine cells inside one NEB carried tdTom labeling with TM exposure at E11.5, suggesting a multi-clonal origin of NEBs. **(C)** TM exposure at E9.5 extensively labeled neuroendocrine cells.

3-6. Early specification of BC-containing epithelia

TM exposure at E9.5 or E8.5 labeled airway epithelial cells in both proximal compartment and distal alveolar compartment. Interestingly, this lineage-tracing result could be consistently recapitulated when TM was administrated at E7.5 and as early as E6.5 (Figures 12A-C). The labeling pattern was not random. When gastrulating embryos at E6.5 were exposed to tamoxifen, only epidermis, a small subpopulation of neuronal cells, laryngeal epithelium, which are known to have p63$^+$ cells in early developmental stages, bear the lineage label; in contrast, heart, liver, kidney and somite, which never have p63 expression, never got labeled (Figure 12A). Interestingly, lung epithelium showed lineage label in all embryos examined, suggesting a potential lung or foregut progenitors specified early in gastrulation stage. This was further confirmed in *p63-CreERT2; Rosa26-lacZ* lineage tracing. The foregut endoderm was extensively labeled at E8.5 when tamoxifen was administered at E6.5 (Figure 12D). When foreguts were dissected from E8.5 embryos exposed to tamoxifen at E6.5 and cultured for another 3 days, similar labeling pattern in the primordial lung bud could be observed (Figure 12E).

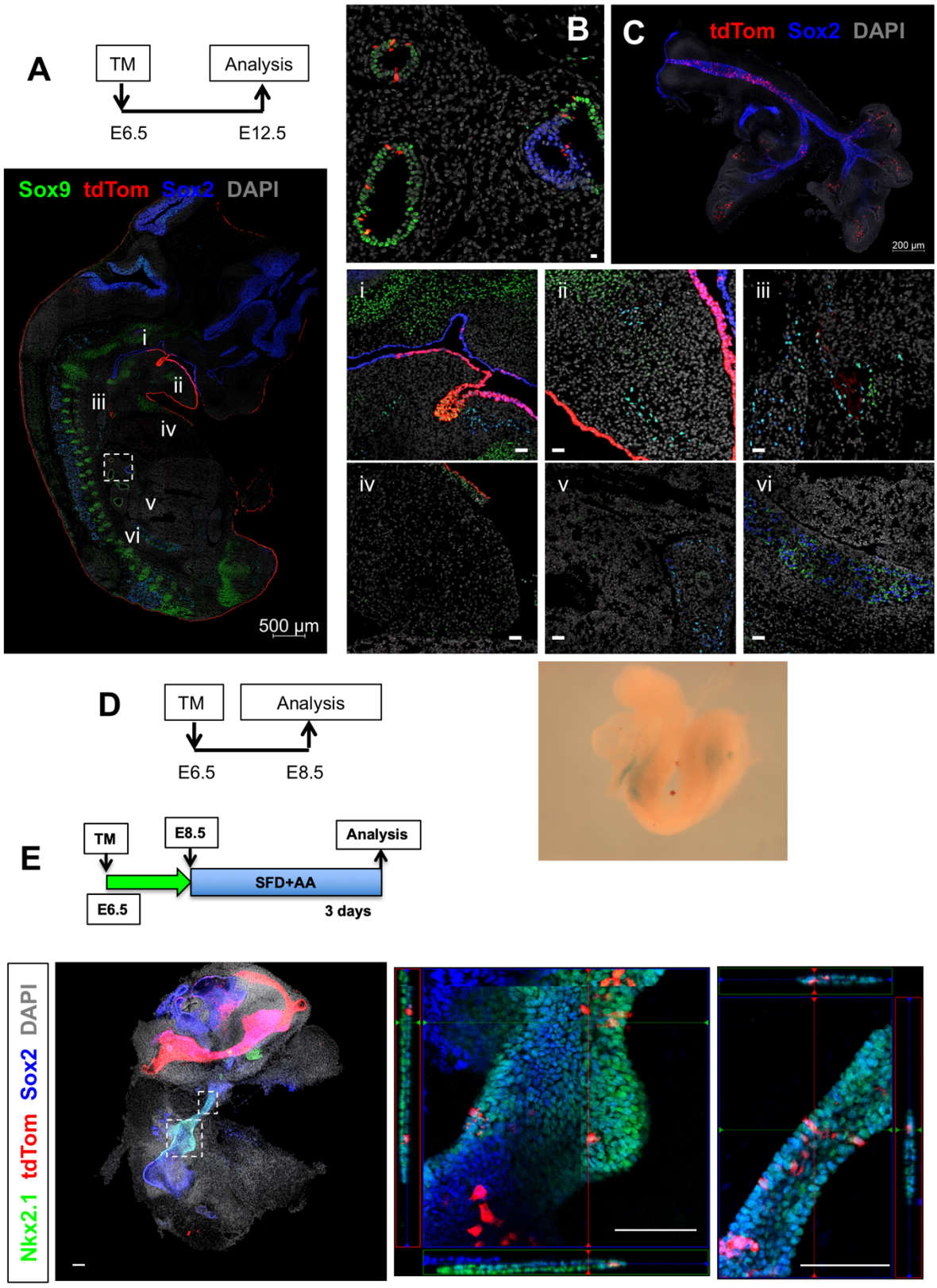


Figure 12. Early labeling of lung epithelium with TM treatment before E8.5.

(A-C) Lineage labeling of *p63-CreERT2; R26-tdTomato* embryos with TM exposure at E6.5 and analyzed at E12.5. **(A)** IF of whole-embryo section showed extensive labeling in in laryngeal epithelium (i), skin (ii), subpopulation of neurons (iii), but never in heart (iv), liver (v), kidney (vi) or somites. **(B)** IF of lung section showed significant labeling in alveolar compartment as well as in the proximal airways. **(C)** Whole-mount immunofluorescence of E12.5 lung showing lineage-labeling in both $Sox2^{+}$ proximal airways and $Sox2^{-}$ distal compartment. **(D)** Lineage labeling of *p63-CreERT2; R26-lacZ* embryos with TM exposure at E6.5 and analyzed at E8.5. LacZ staining showed extensive labeling of p63 lineage in foregut endoderm. **(E)** *p63-CreERT2; R26-tdTomato* embryos were exposed to TM at E6.5, foreguts were then isolated at E8.5 and cultured for 3 days. IF of foregut explants after 3 days in culture showing that the primordial lung buds were labeled by tdTomato. The lineage-labeling patterns in both cases were similar to the *in vivo* tracing data. Scale bars: 50μm in **(A & B)** 100μm in **(E)**.

Chapter 4. Dual roles of p63 in airway development

4-1. p63 is required for the specification of airway BCs, but dispensable for luminal lineage balance

The lineage tracing analyses suggested that embryonic $p63^+$ progenitors made equal contributions to multiciliated versus secretory lineages in luminal compartment during embryonic development. Previous publications suggested that knockout of p63 gene resulted in a disrupted balance in luminal lineages, in which multiciliated cells increased significantly at the expense of secretory cells[82,118]. As the several p63-deficient mice generated contradictory results, there has been no definitive conclusions for more than a decade[66,68-70,77,78,119,120]. To solve this issue, the *p63-CreERT2* knock-in allele was used to generate a new homozygous deletion of p63 knockout model.

This knockout (KO) model could be more suitable for correlations with the findings of the lineage analysis. The KO animals phenocopied the severe skin and limb defects previously reported[68,77]. The tracheal epithelium showed no $p63^+$ $Krt5^+$ cells (Figures 13A & B). Similar expression of CC10, Scgb3a2, and Foxj1 in E18.5 wild-type (WT) and KO tracheas by immunofluorescence and qRT-PCR suggested that p63 did not influence secretory versus multiciliated fate specification (Figures 11A & B). This was in agreement with the comparable distribution of tdTom in secretory and multiciliated cells in lineage-tracing analysis (Figure 8A). No difference in intrapulmonary airways including neuroendocrine lineage was observed (Figure 13D), confirming the minimal effect of p63 on intrapulmonary airway cell fates.

In contrast to the requirement of p63 to maintain the proliferative potential of stem cells in skin and thymus[71], the proliferative activity of airway epithelium was not affected in the *p63CreERT2/CreERT2* knockout animals (Figure 14A), consistent with the observation that $p63^+$ cells did not exhibit higher proliferative index in the EdU assay compared to the rest epithelial cells in the above three lineage tracings (Figure 7A).

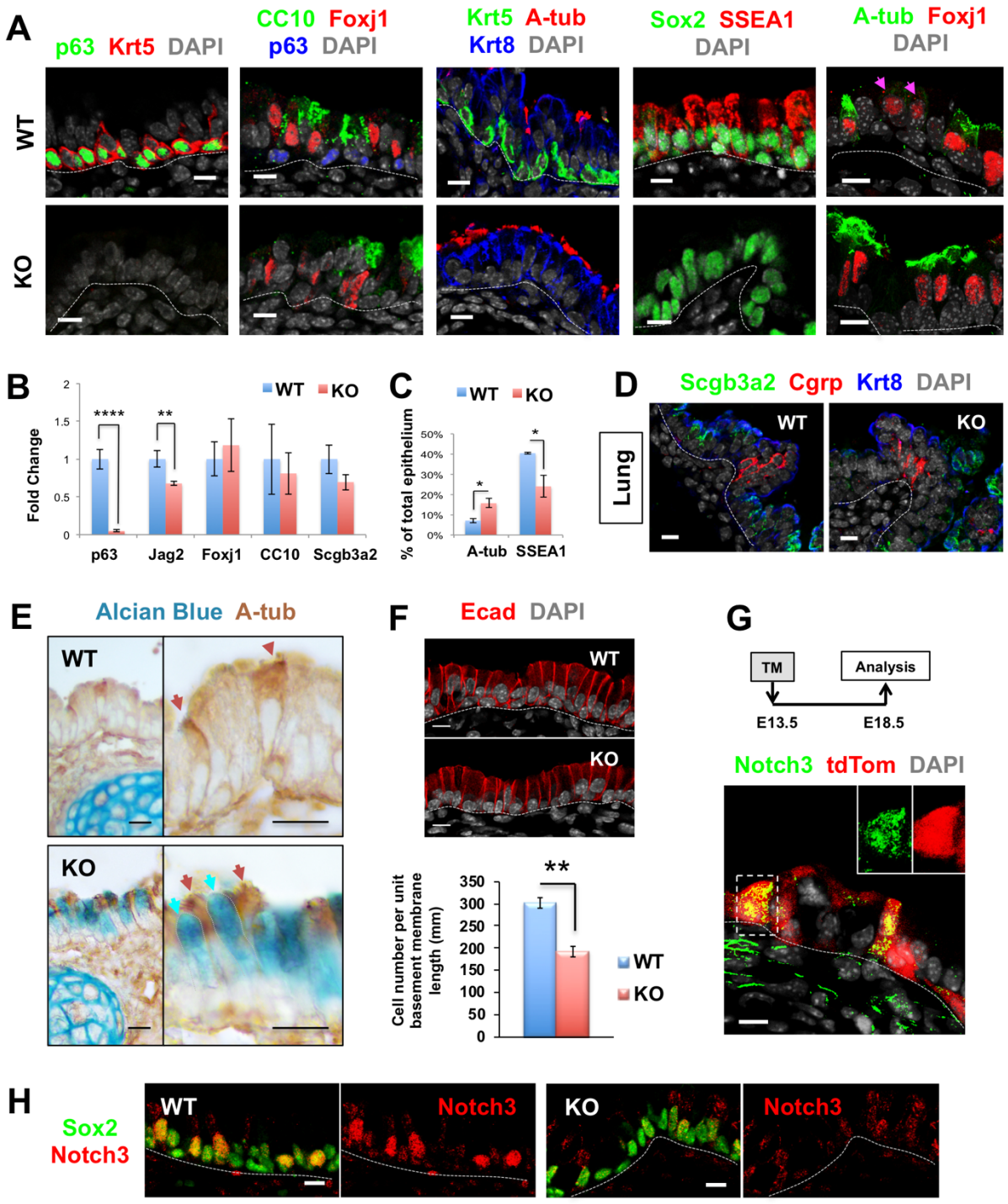


Figure 13. p63 KO mutants have preserved luminal lineage balance with altered maturation and organization of the pseudostratified epithelium.

(A-E) Tracheal differentiation in E18.5 WT and p63 null (KO) mice. **(A)** IF, markers of basal (p63, Krt5), luminal (Krt8), secretory (CC10, SSEA1), multiciliated (Foxj1, A-tub), airway progenitor (Sox2) cells. Far right panel: maximum projection of confocal z-stack images showing colocalization of A-tub$^+$ cilia and Foxj1$^+$ nuclei; magenta arrows showing Foxj1$^+$ multiciliated cells without cilia. **(B)** qRT-PCR, *p63, Foxj1, CC10, Jag2* (Notch ligand), *Scgb3a2* (secretory marker), **(C)** Morphometric analysis of A-tub and SSEA1. (**D)** E18.5 intrapulmonary airways of WT and p63 KO showing similar expression of Cgrp and Sgb3a2 (neuroendocrine and secretory cell markers). **(E)** Alcian Blue and A-tub staining of E18.5 WT and KO tracheas: Brown arrows showing A-tub$^+$ cilia; Cyan arrows showing Alcian Blue$^+$ goblet cells. **(F)** IF showing simplified epithelial structure from the pseudostratified tracheal epithelium in the KOs at E18.5; graph: epithelial cell number per unit basement membrane length (mm). **(G)** *p63-CreERT2* lineage tracing and IF showing Notch3$^+$ tdTom$^+$ in tracheal epithelium at E18.5 (TM administration at E13.5); insets: single channels depicted from boxed area. **(H)** IF E18.5 tracheal sections: Strong Notch3 nuclear staining in suprabasal cells of WT (nuclear Sox2 in all epithelial cells), however in p63 KO Notch3 was diffusely expressed at low levels in cytoplasm and no longer in the nucleus. Graphs: mean ± SEM from 6-12 fields per sample, N$\geqslant$3. Statistics: Student's t-test: *P<0.05; **P<0.01; ****P<0.0001; n.s. non-significant. Scale bars: 10μm.

4-2. p63 is required to prevent advanced maturation in luminal progenitors

Although no alteration in the balance of two major luminal lineages was observed in the p63 null animals, A-tub labeling was more abundant in KO epithelia at E18.5, suggesting that once committed, ciliated cells matured to a greater extent to form multicilia (Figures 13A and 13C). In E18.5WT animals, not all Foxj1$^+$ ciliated cells had A-tub$^+$ multicilia. Nevertheless, the Foxj1 A-tub double-positive multiciliated cells were overrepresented in the p63 KO trachea. This suggested that even though the specification of ciliated lineage was not affected, after cell fate selection, the maturation process of multiciliated cells was advanced in the p63 KO. This was actually consistent with the reported phenotype in Daniely et al., 2004[82], as they only used tub IV, a late marker for mature cilia detection, as the marker to evaluate abundance of multiciliated cells.

Moreover, expression of SSEA1, which marks immature secretory cells[121], was greatly reduced in KO tracheas both at early (E15.5) and late (E18.5) stages, while unaffected in intrapulmonary airways (reflecting the regional distribution of $p63^+$ cells in WT) (Figures 13A, 13C, 13D, 14B). By contrast, numerous Alcian blue-labeled cells in KO indicated that, without p63, secretory progenitors undergo aberrant goblet cell differentiation (Figure 13E).

As expected from the absence of BCs, the pseudostratified architecture of the tracheal epithelium was converted into a columnar epithelium with a reduction in the epithelial cell number by 36.4% in KOs (Figure 13F). Nuclear $Notch3^+$ $tdTom^+$ was present in E18.5 *p63-CreERT2* tracheas (Figure 13G; Mori et al., 2015[122]). However, in KO tracheas Notch3 was expressed diffusely and no longer in the nucleus, likely from the limited Jag ligand available for Notch3 activation without $p63^+$ cells (Figure 13H). Indeed qRT-PCR showed decreased Jag2 expression (Figure 13B). Low Jag levels, presumably from multiciliated cells, could still induce expression but not activation of Notch3. Altered Notch3 could potentially influence differentiation behavior but not fate choice. Thus, the phenotype observed could be potentially explained by the loss of BC progenitors in the absence of p63, with an alternative progenitor (p63-independent) then undergoing an aberrant luminal differentiation program. This would suggest that prenatal p63 is crucial in generating not only the postnatal BC pool, but also the embryonic progenitors of luminal cells able to carry the normal differentiation of the developing airways.

Laryngeal epithelium is continuously extended from tracheal epithelium, but these two epithelia exhibit important differences. In contrast to the pseudostratified architecture of tracheal epithelium, laryngeal epithelium is stratified, similar to esophagus and skin. Cells in laryngeal epithelium are negative for Nkx2-1 but positive for Sox2, representing an endodermal lineage. Similar to other stratified epithelia like esophagus and skin, laryngeal epithelium is composed of a basal-cell layer with luminal cells on top. In p63 KO animals, the E15.5 laryngeal epithelium exhibited a more advanced maturation compared to the tracheal epithelium. As expected, the stratified architecture was disrupted in the p63 null animals. Interestingly, many luminal cells in E15.5 WT laryngeal epithelium showed cytoplasmic immunostaining signal for A-tub, suggesting an intracellular organization prepared for the initiation of ciliogenesis. In E15.5 KO animals, more ciliated cells with multicilia were identified (Figure 14C). Moreover, the SSEA1 expression in luminal compartment was almost abolished in p63 KO laryngeal epithelium, reminiscent of

the E15.5 tracheal phenotype (Figure 14C). The difference in phenotype severity between trachea and larynx can be explained by different developmental mechanisms adopted in pseudostratified epithelium versus stratified one. In pseudostratified epithelium like trachea, the cells at early developmental stages (before E15.5) broadly express p63 at various and relatively low levels. Through some currently unknown signaling pathways (probably Notch due to lateral inhibition), some cells upregulate p63 and acquire BC fate, while others downregulate p63 and adopt luminal fate. This basal-luminal segregation was described above in the lineage-tracing analyses. In contrast, stratified epithelium like larynx initially starts with one single layer of p63-expressing BCs. These cells proliferate and give rise to luminal descendants on top during the switch from simple columnar epithelium to stratified architecture. So, it is possible that whether luminal descendants need to pass through a BC status affects the final outcome of their differentiation status.

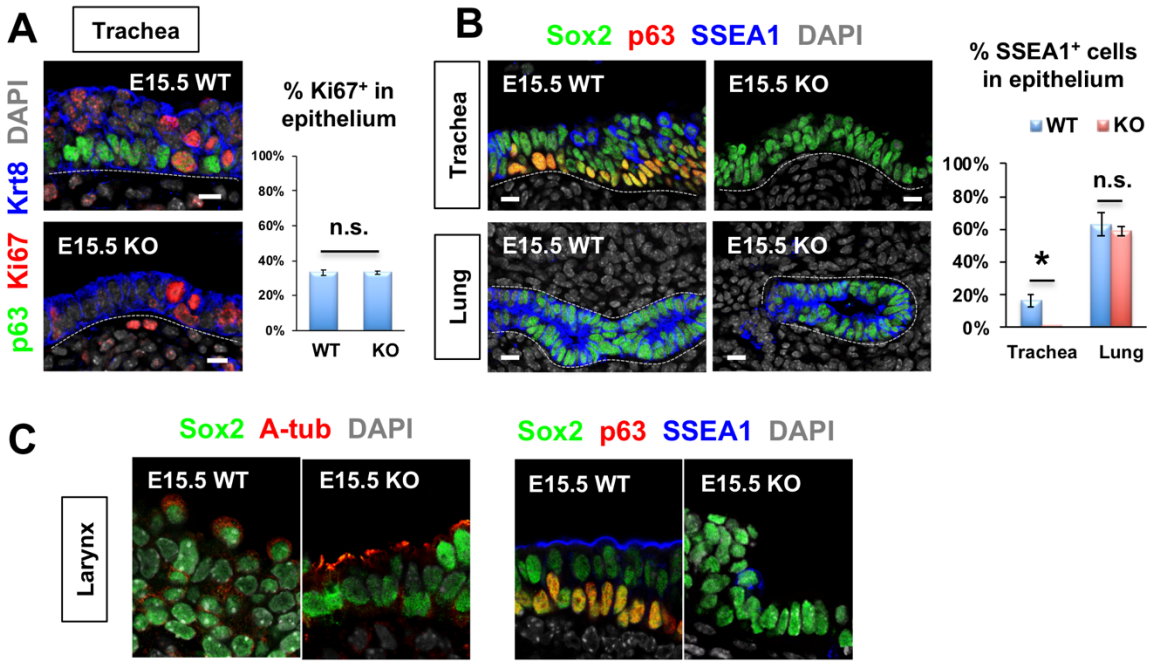


Figure 14. p63 KO animals exhibited advanced maturation phenotype as early as E15.5 before the appearance of characteristic BCs.

(A) IF of E15.5 tracheas showing no difference in proliferation index (Ki67) in epithelium between WT and p63 KO mice. **(B)** IF at E15.5: marked reduction in SSEA1-expressing cells in KO selectively in trachea not

in intrapulmonary airways (lung; note Sox2 epithelial labeling in all panels but p63 restricted in WT). **(C)** IF of E15.5 larynx showing advanced ciliogenesis and marked reduced SSEA1 in p63 KO mice. Graphs in **(A-B)** are mean ± SEM from 6-12 fields per sample, N⩾3. Student's t-test: *P<0.05; n.s. non-significant. Scale bars: 10μm.

Chapter 5. Embryonic origin of bronchial progenitors responsible for H1N1-induced Krt5$^+$ pods

5-1. Adult p63$^+$ Krt5$^-$ bronchial progenitors generating Krt5$^+$ pods after H1N1 challenge

In adult murine airway system, p63$^+$ Krt5$^+$ BCs are restricted to trachea and extrapulmonary main stem bronchi. However, in extreme scenario that extensive damage is caused by H1N1 influenza virus in mouse lung, the ectopic Krt5$^+$ p63$^+$ basal-like cells emerge in intrapulmonary airways (airway Krt5$^+$ cells) and cluster into "Krt5$^+$ pods" in interstitial space (alveolar Krt5$^+$ cells)[101,102,105,106].

To investigate the contribution from p63 lineage to the Krt5$^+$ pods induced by H1N1 influenza infection, *p63-CreERT2; R26-tdTomato* adult mice were administered tamoxifen and infected with PR8 virus after 21 days (Figure 15A). Analysis of these lungs showed massive expansion of the Krt5$^+$ p63$^+$ pods in the lung interstitium. These pods were nearly always colocalized with tdTomato labeling (Figures 15B-E). Quantification showed that more than 99.5% ± 0.5% and 98.9% ± 1.0% of airway and alveolar Krt5$^+$ cells (N=5 from 3 independent experiments), respectively, were p63 lineage-labeled (Figure 15F). Notably, the Krt5$^+$ pods were preferentially located in alveolar regions closely associated with main stem bronchi (Figures 16A-C; Zuo et al., 2015[102]). This suggested that the progenitors of Krt5$^+$ pods could have originated from main stem bronchi.

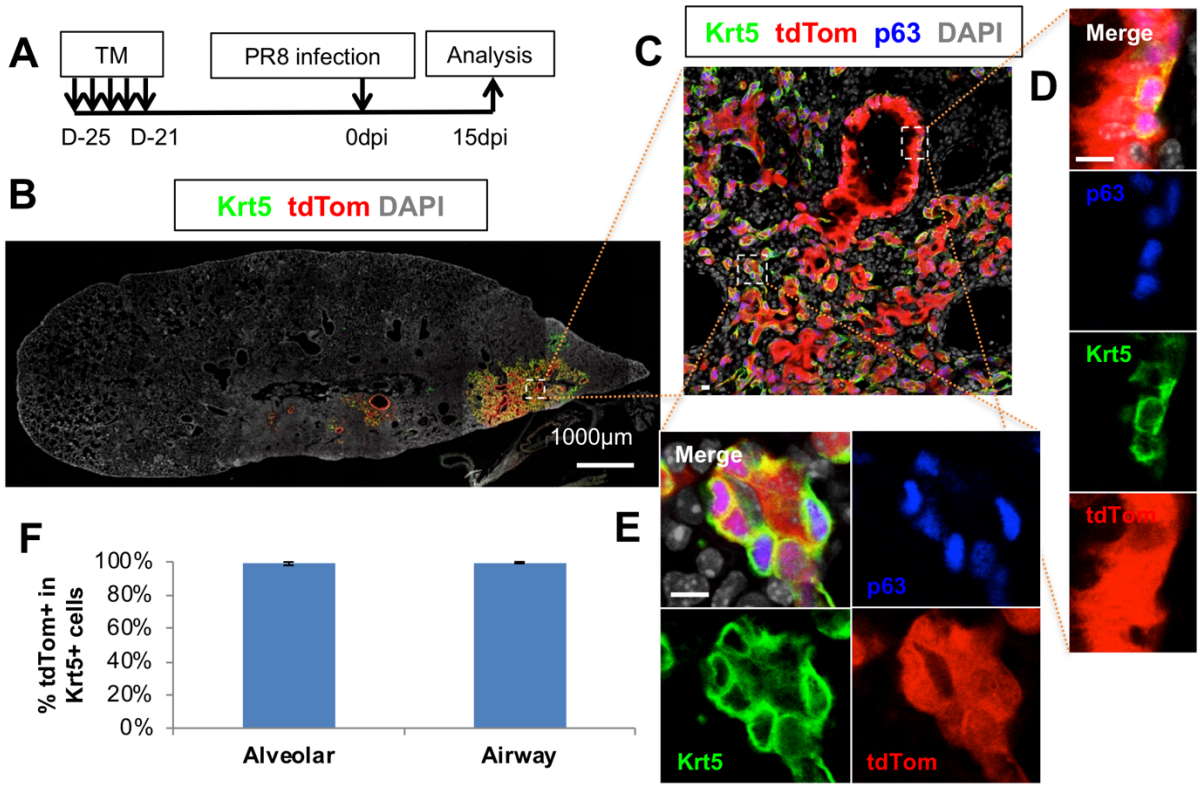


Figure 15. p63 lineage-labeled cells generating the Krt5⁺ pods after H1N1 challenge.

Lineage labeling of *p63-CreERT2; R26-tdTomato* mice infected with H1N1-PR8 21 days after TM administration in adulthood. **(A-F)** IF analysis of 15dpi lungs showing continuous trail of tdTom⁺ cells from large airway to alveolar space with nearly 99% tdTom⁺ labeling of the ectopic Krt5⁺ cells in the airway **(C & D)** and alveolar **(C & E)** compartments. Graph: mean ± SEM from 10-12 fields per sample. n=5 from 3 independent experiments. Scale bars: 10μm unless noted.

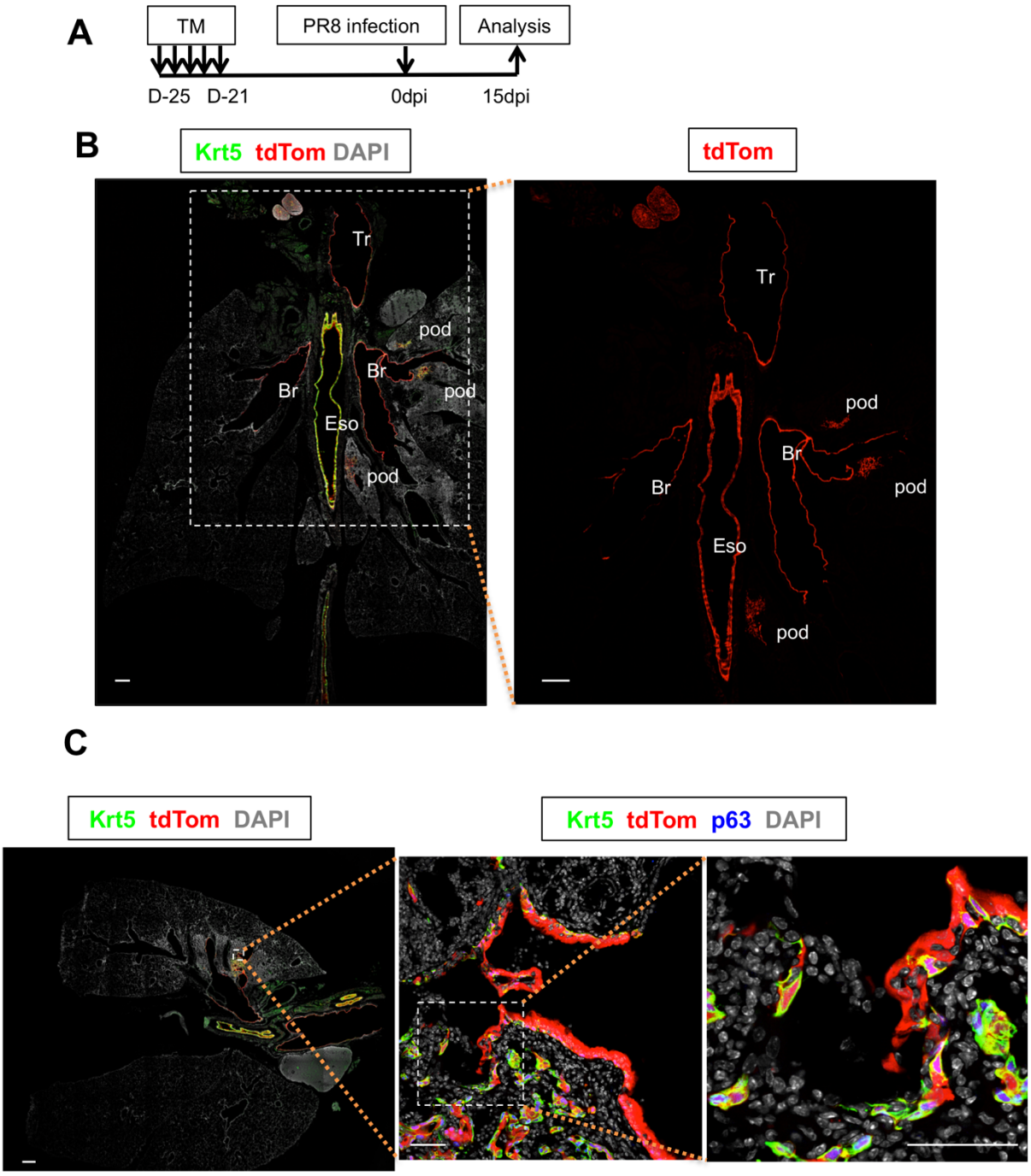


Figure 16. H1N1-induced Krt5^{+} pods radiated outward from the bronchi.

(A) Animals exposed to 5 doses of tamoxifen in adulthood and infected with H1N1 21 days after the last dose of tamoxifen to ensure no residual effect. **(B)** Tile scan IF of 15dpi lung section showing that lineage-labeled Krt5^{+} pods were closely adjacent to lineage-labeled mainstem bronchi. **(C)** IF of 15dpi lung section showing that lineage-labeled cells invading interstitial space from proximal airways. Scale bars: 50μm.

5-2. Time course analysis of H1N1 injury response

To closely examine the cellular origin of these Krt5^{+} pods and their potential migrating trail, a sequential analysis of their behavior was conducted during H1N1-induced injury. At 0dpi, several p63^{+} Krt5^{-} cells bearing lineage-label were seen at the boundary between intrapulmonary and extrapulmonary main stem bronchi (Figures 17A-D). At 8dpi these cells expanded toward adjacent regions in lobar bronchi (Figures 17E-G). At 11dpi, a few Krt5^{+} cells started to appear in the p63^{+} lineage-labeled epithelium in proximal airways. Moreover, in areas of more severe injury, small pod-like Krt5^{+} p63^{+} tdTom^{+} structures were first observed (Figures 17H-K). Notably the most distally located lineage-labeled p63^{+} cells assumed the basal location, suggestive of a migratory behavior (Figure 17I-K). A trail of p63-lineage-labeled cells was more apparent at 15dpi and in association with adjacent interstitial Krt5^{+} pods (Figures 16B-C, 17L). Consistent with previous reports (Vaughan et al. 2015), these Krt5^{+} pods failed to resolve and persisted as Krt5-expressing cysts at 58dpi (Figure 17L). The stepwise acquisition of Krt5 and their noticeable trail from large airways (where we identified p63^{+} Krt5^{-} tdTom^{+} cells in uninfected lungs) to areas undergoing remodeling, support their short-range migratory behavior and the idea that progenitors for these lesions reside in the main bronchi.

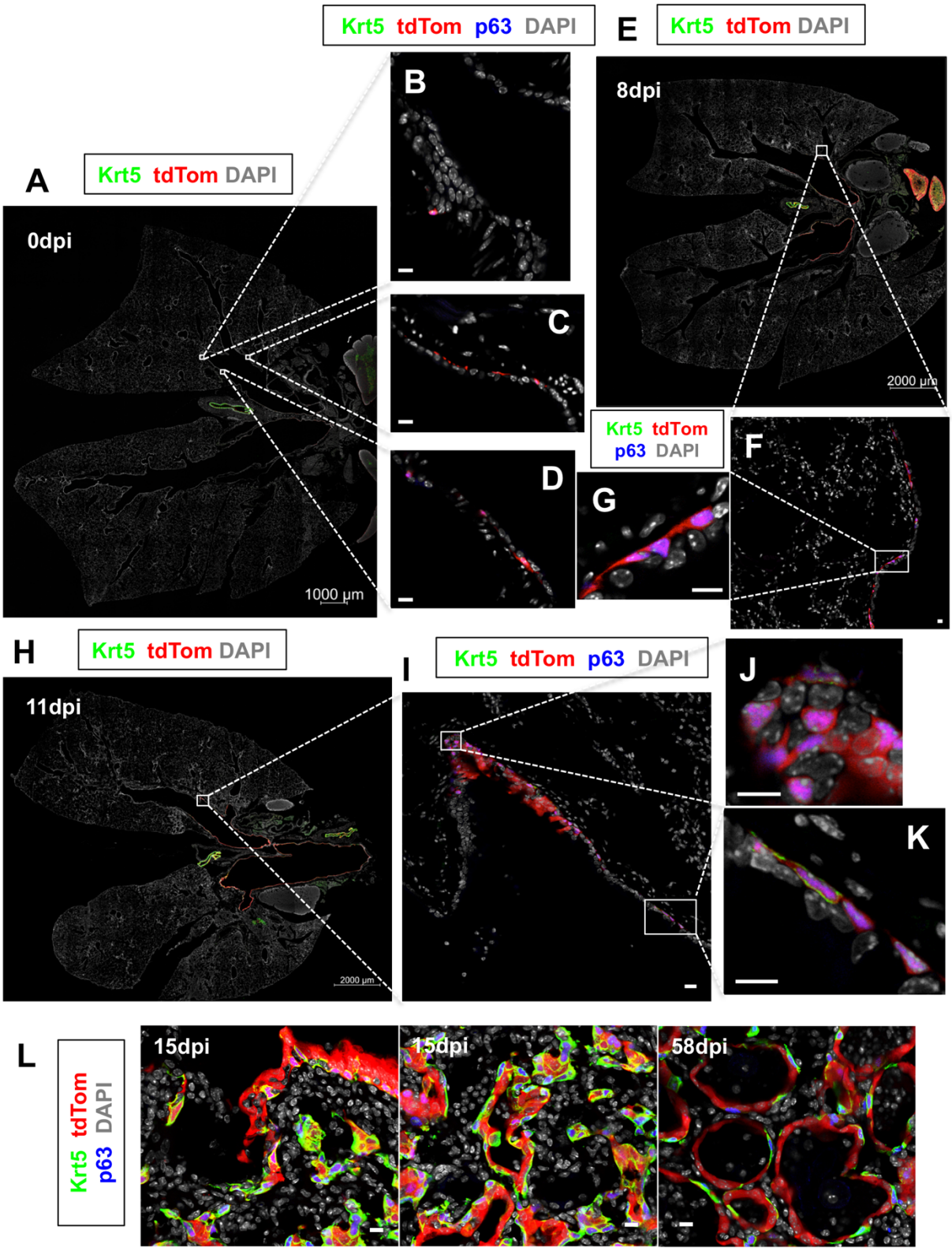

A
Krt5 tdTom DAPI
0dpi
1000 µm
B
Krt5 tdTom p63 DAPI
C
D
E
Krt5 tdTom DAPI
8dpi
2000 µm
F
Krt5 tdTom
p63 DAPI
G
H
Krt5 tdTom DAPI
11dpi
2000 µm
I
Krt5 tdTom p63 DAPI
J
K
L
Krt5 tdTom
p63 DAPI
15dpi
15dpi
58dpi

Figure 17. Time course study indicated that the p63 lineage-labeled resident p63$^+$ Krt5$^-$ progenitors in main bronchi migrated along proximal airways towards the area with most severe damage to generate Krt5$^+$ pods after H1N1 challenge.

(A, E, H) Tile scan of lung IF showing the proximal-to-distal progression of p63 lineage-labeled cells in main stem bronchi at 0dpi, 8dpi and 11dpi. **(B-D)** tdTom-labeled p63$^+$ Krt5$^-$ cells showed in high magnification residing in main stem bronchi near lung boundary at 0dpi. **(F, G, I, J, K)** More tdTom-labeled p63$^+$ Krt5$^-$ cells can be identified at basal location of intrapulmonary airways at 8dpi **(F, G)** and at 11dpi **(I, J, K)**. **(L)** Extensive tdTom labeling of the airway epithelium with migrating trail towards adjacent alveolar space to form the p63$^+$ Krt5$^+$ tdTom$^+$ clusters and honeycomb-like structures (15dpi) which persisted by 58dpi. Scale bars: 10μm unless noted.

In PBS-mock infection, no lineage-labeled trails have been identified along intrapulmonary main stem bronchi (Figure 18A). Krt5$^-$ p63$^+$ bronchial progenitors could be identified in the intrapulmonary main stem bronchi, but these cells did not expand without injury. To evaluate the effect from TM-independent recombination events, *p63-CreERT2; R26-tdTomato* mice without exposure to TM were infected with H1N1 and examined at 15dpi. Although rare tdTom$^+$ cells could be identified in the H1N1-induced Krt5$^+$ pods at 15dpi, the lineage labeling contribution was minimal (Figure 18B). Therefore, the p63$^+$ Krt5$^-$ bronchial progenitors residing in the most proximal part of large intrapulmonary airways are the only cellular source for Krt5$^+$ pods after H1N1 challenge.

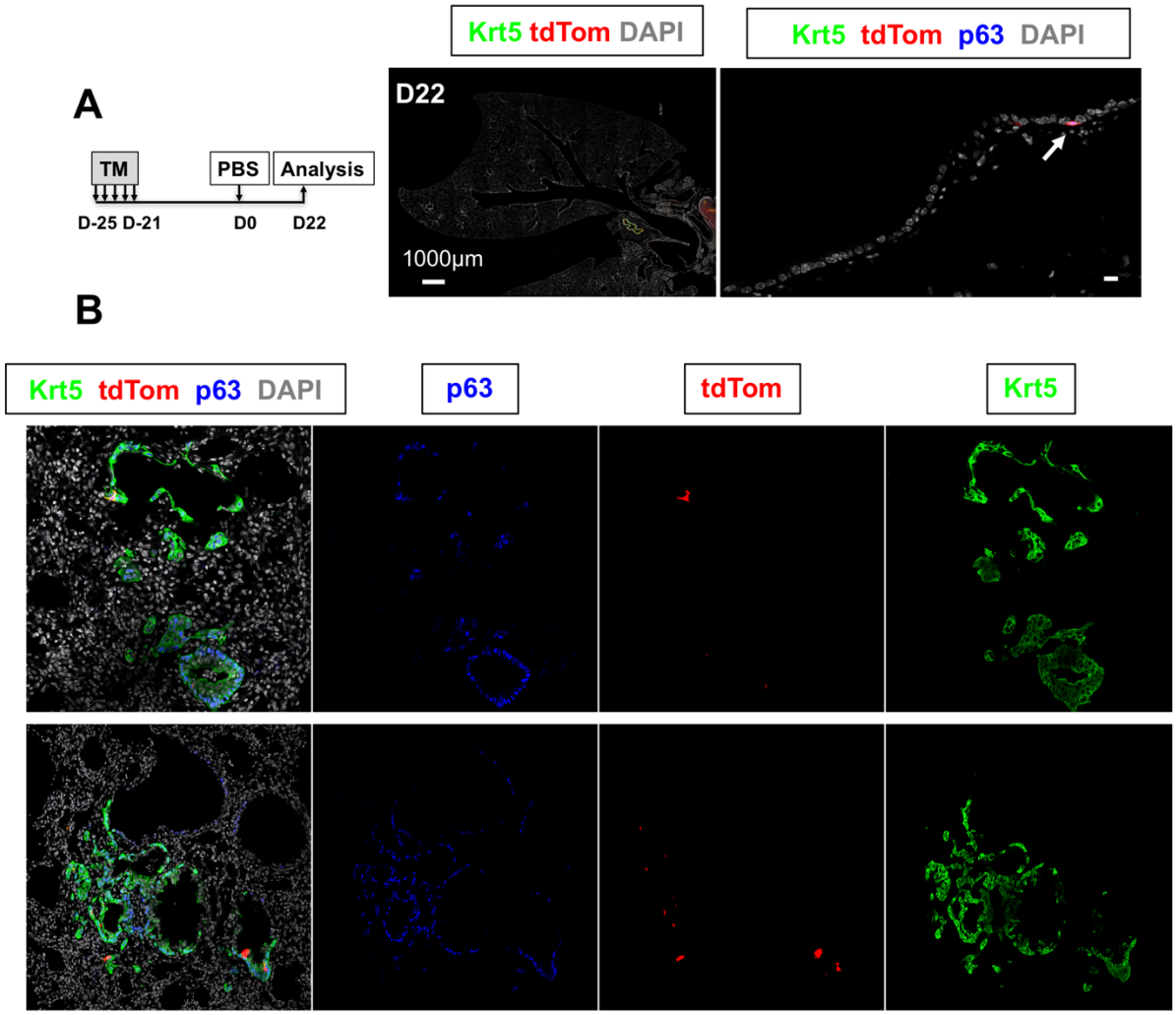


Figure 18. Generation of Krt5^{+} pods depended on H1N1 challenge, and TM-independent recombination events contributed minimally to the lineage labeling.

(A) PBS mock-infection showing the rare lineage-labeled p63^{+} Krt5^{-} intrapulmonary progenitors (arrow) unaltered after 22 days. **(B)** IF analysis of 15dpi lung sections from *p63-CreERT2; R26-tdTomato* mice without TM exposure showing minimal tdTom labeling in the H1N1-induced Krt5^{+} pods. Scale bars: 10μm unless noted.

5-3. Persistence of injury-induced Krt5^{+} pods

The ectopic appearance of p63^{+} Krt5^{+} basal-like cells in interstitial lung after H1N1 injury is intriguing and has motivated extensive debate whether these cells are adult stem cells responsible for alveolar

regeneration[101,102,105-107]. To solve this controversy, the lineage commitment of H1N1-induced Krt5$^+$ pods was examined for an extended time after virus exposure. *p63-CreERT2; R26-tdTomato* system with TM administration in adulthood and analyzed at 58dpi showed lineage labeling in almost 100% of the Krt5$^+$ pods (Figure 19A). Consistent with Ray, et al., 2016[106], Vaughan, et al., 2015[105], Xi, et al., 2017[107], the lineage-labeled cells were found to persist as cysts in the damaged lung regions, reminiscent of the Krt5$^+$ pathological scars in IPF (Figure 19B; Vaughan, et al., 2015[105]). Comparing to the Krt5$^+$ pods at 15dpi, the cysts were dilated with bigger lumen and Krt5$^+$ cells seemed to switch from the squamous morphology to a relatively flattened cellular structure at 58dpi (Figure 19B). A considerable amount of lineage-labeled cells no longer expressed Krt5 or p63, and mucus could be observed in the lumen, suggesting a putative bronchiolization process. Lineage-labeled cells did express Pdpn, a marker for AT1 cells (Figure 19C). However, Pdpn is also expressed in airway BCs, rendering the conclusion that Krt5$^+$ pods could resolve into alveolar lineages unsupported. Almost no pro-Spc$^+$ tdTom$^+$ cells were observed (Figure 19C), suggesting limited contribution to AT2 lineage. In extremely rare cases, tdTom$^+$ cells co-expressing Pdpn and pro-Spc could be identified, which might represent some limited cellular plasticity of these basal-like cells towards AT2 lineage.

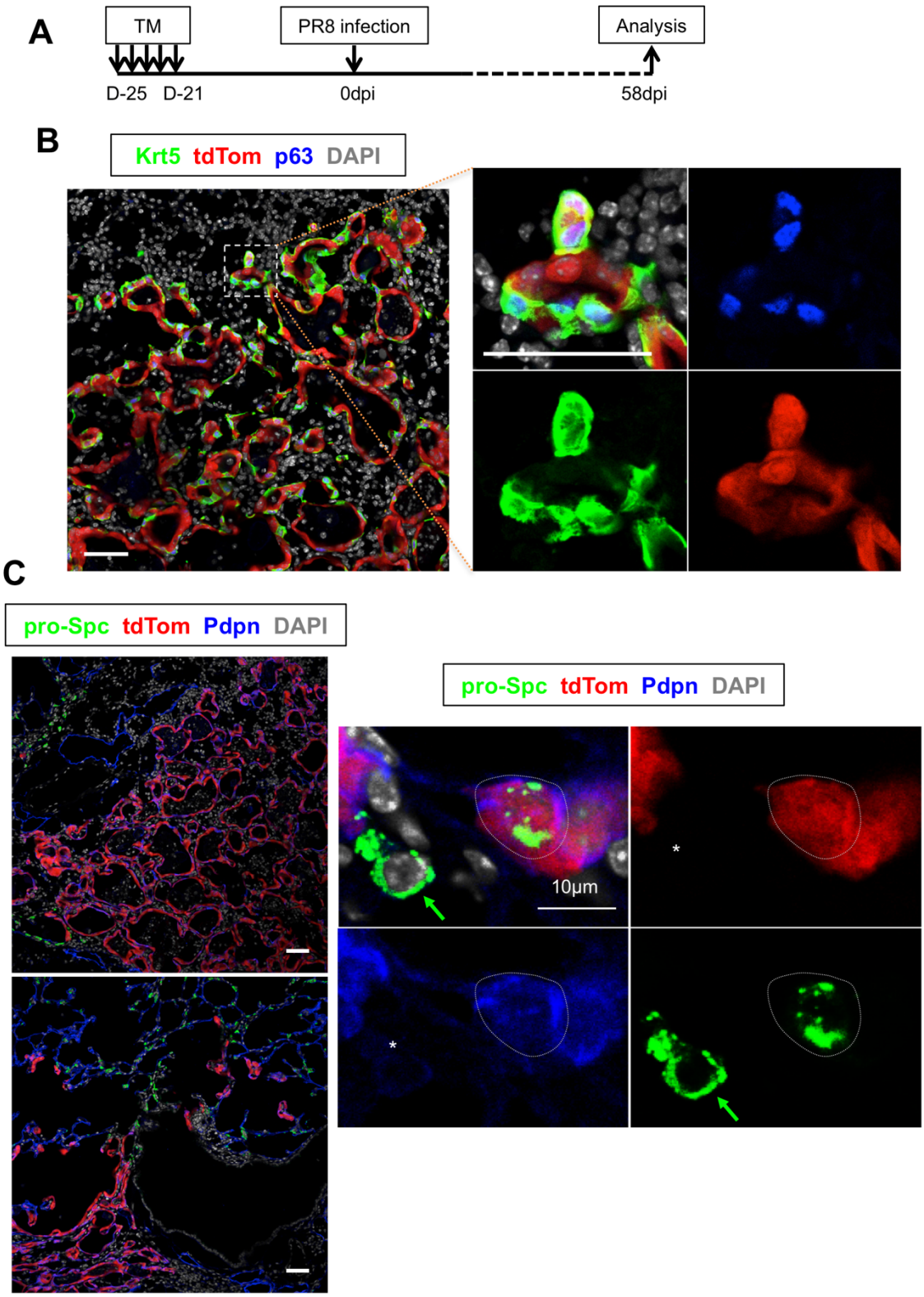


Figure 19. H1N1-induced Krt5^{+} pods persist with minimal regenerating activity at 58dpi.

(A) Animals exposed to 5 doses of tamoxifen in adulthood, infected with H1N1-PR8 influenza virus and analyzed at 58dpi. **(B)** IF of lung section showing lineage-labeled Krt5$^+$ p63$^+$ pods persisted in honeycombing structure. **(C)** IF analysis of 58dpi lungs showing persisted Pdpn$^+$ tdTom$^+$ cysts with minimal contribution to alveolar regeneration. *no fluorescent signal. Scale bars: 50μm unless noted.

The lineage-tracing analyses showed that before E10.5, p63$^+$ progenitors are multipotent to generate both airway and alveolar lineages, but this potential was lost by E10.5. The results in the H1N1 injured adult lungs showed that not even in the presence of catastrophic damage, the potential to generate alveolar fates is negligible. It is possible that epigenetic modification at E10.5 permanently restricted the plasticity of p63$^+$ cells. Future investigations will be of great scientific significance to identify the molecular mechanisms for this fate restriction and to steer the regenerative potentials towards alveolar repair.

5-4. Embryonic p63$^+$ progenitors as the developmental origin of Krt5$^+$ pod precursors

The lineage-tracing analyses described here showed that a subpopulation of p63$^+$ progenitors at E13.5 was maintained undifferentiated and established the basal stem cell pool of the adult trachea. To investigate whether this was also true in intrapulmonary airways, E14.5 lung sections were closely examined, revealing scattered p63$^+$ cells in the intrapulmonary main stem bronchi with Krt8 expression (Figure 20A). The immunostaining signal of p63 in these intrapulmonary population was remarkably weaker than those in extrapulmonary airways (bona fide airway BCs). To follow the fate of these embryonic intrapulmonary p63$^+$ progenitors, the same lineage tracing regimen was used with TM administration at E13.5. By E18.5, most of the tdTom$^+$ cells were co-labeled with multiciliated cell marker Foxj1 or secretory cell marker CC10, or neuroendocrine marker CGRP (Figure 20B), suggesting that they contribute to the pool of differentiated epithelial cells during airway development. Again, consistent with previous analysis, no pro-Spc$^+$ AT2 cells were labeled in this E13.5 tracing. Although not all intrapulmonary p63$^+$ cells got labeled at E14.5 (Figure 21A), tdTom$^+$ p63$^+$ Krt5$^-$ cells were present in E18.5 lungs (Figure 21B), suggesting that a subpopulation of these cells remained uncommitted by late gestation stage. Interestingly, these cells never acquired Krt5 and remained immature throughout lifetime (Figure 21C).

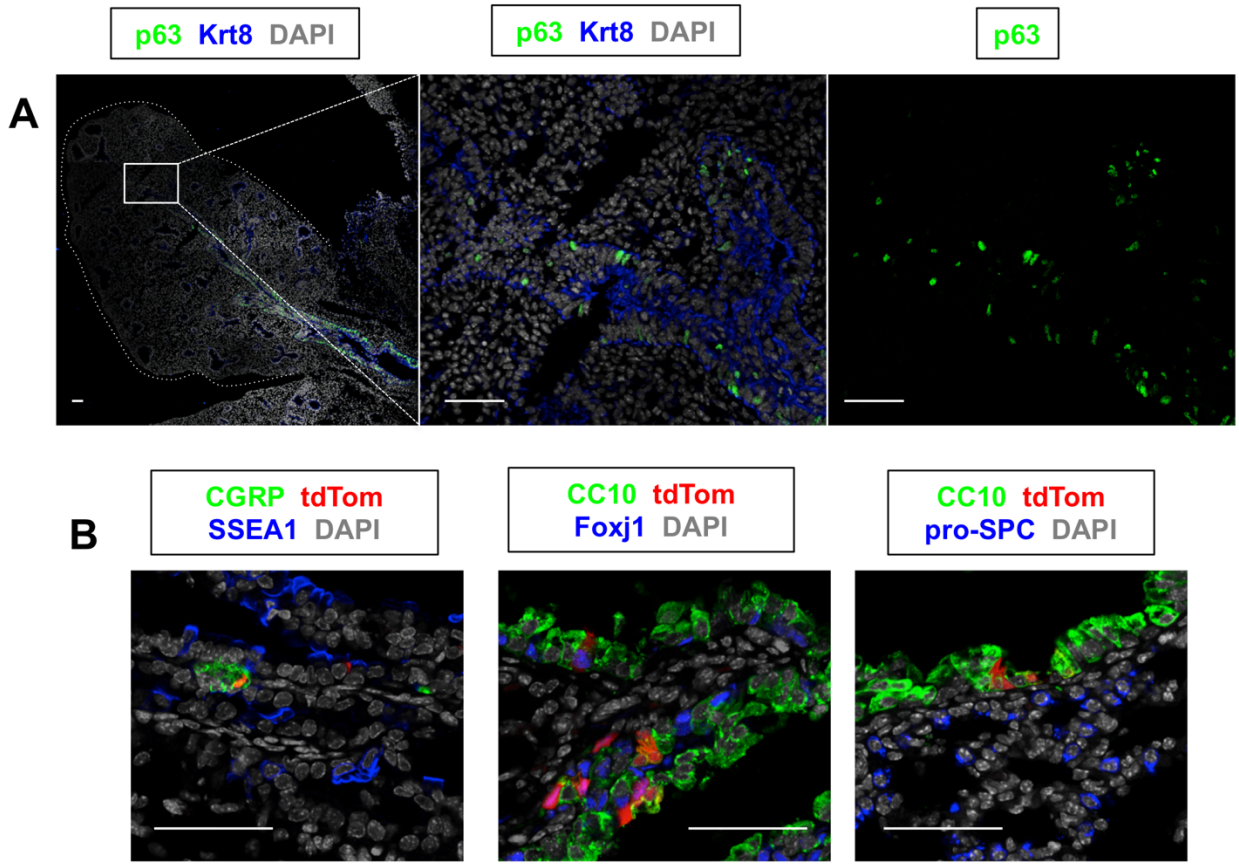


Figure 20. Embryonic intrapulmonary p63⁺ cells could differentiate into all luminal lineages in proximal airways.

(A) IF of E14.5 lung section showing scattered p63⁺ cells in intrapulmonary main bronchi. **(B)** IF tdTom double-labeled with markers of airway/alveolar differentiation. Scale bars: 50μm.

Morphometric analyses showed that recombination efficiency in intrapulmonary p63⁺ cells in this E13.5 tracing was 6.7% ± 1.4%, markedly lower than the 35.7% ± 4.5% in tracheal p63⁺ cells (Figure 21D). This reflected the major difference in p63 promoter activities in these regionally distinct p63⁺ populations. Interestingly, with such low labeling efficiency and the fact that more than 96% of these lineage-labeled cells committed to luminal differentiation by E18.5, 53.9% ± 0.9% of the intrapulmonary p63⁺ Krt5⁻ cells bearing the tdTom label at E18.5, and 43.1% ± 21.6% in adulthood could still be detected (Figure 21D).

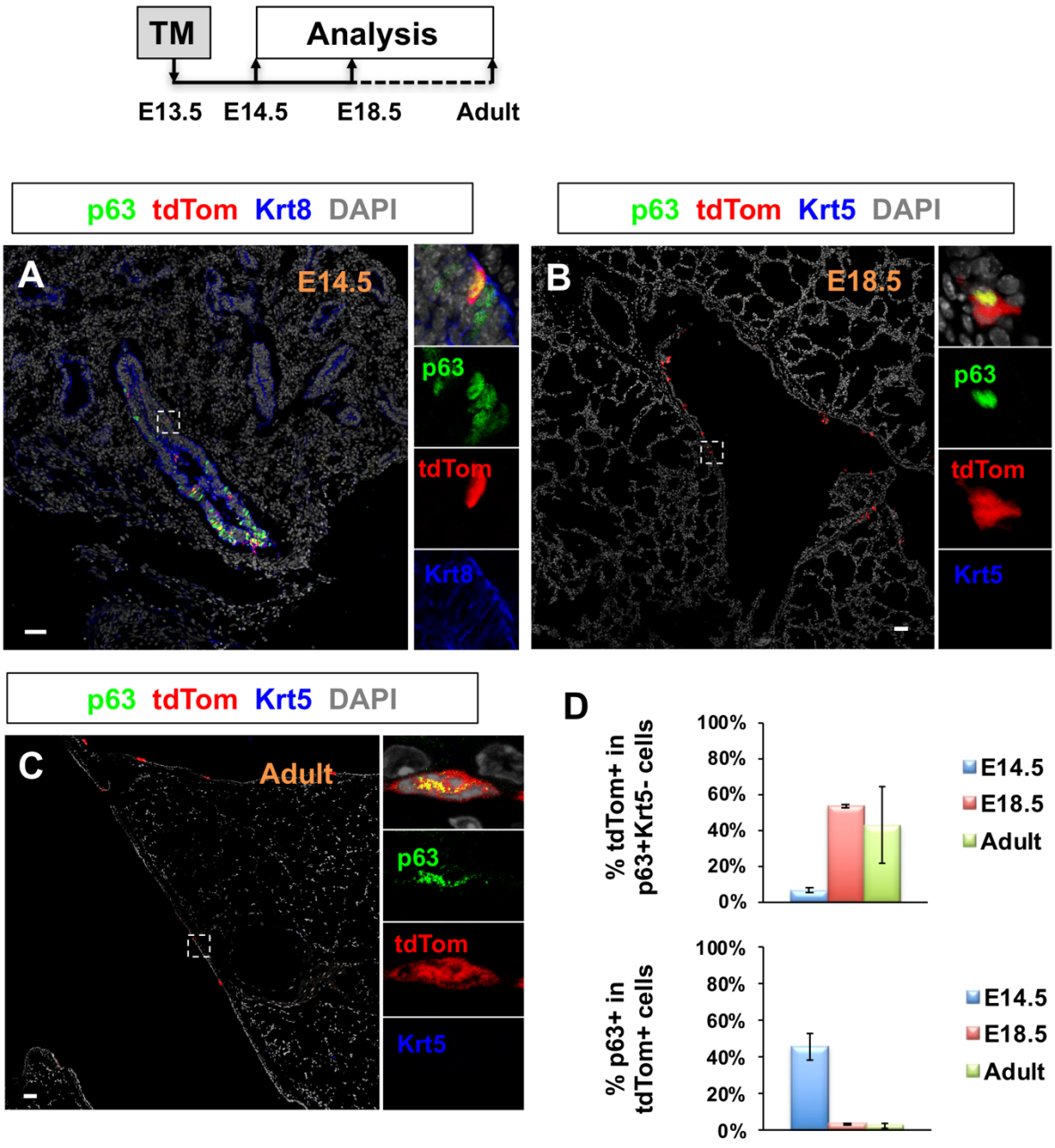


Figure 21. Embryonic intrapulmonary p63⁺ progenitors are maintained immature throughout adulthood.

(A-C) Lineage tracing of p63⁺ descendants in intrapulmonary airways at E14.5, E18.5 and adult lung: TM in E13.5 *p63-CreERT2; R26-tdTomato* mice. IF tdTom double-labeled with p63 **(A-C)**. **(D)** Top graph: % tdTom⁺ p63⁺ Krt5⁻ cells in total intrapulmonary p63⁺ Krt5⁻ cells; bottom graph: % tdTom⁺ p63⁺ cells in total

intrapulmonary tdTom$^+$ cells. Graphs: mean ± SEM, N⩾3 except E18.5 in E13.5 tracing (N=2). Scale bars: 50μm.

Then, the participation of these intrapulmonary p63$^+$ Krt5$^-$ tdTom$^+$ progenitors labeled prenatally was investigated in the adult lung response to injury. The basal-like morphology and p63 lineage suggested their involvement in the reported H1N1-induced aberrant alveolar remodeling[107]. *p63-CreERT2; R26-tdTomato* embryos exposed to TM at E13.5 were infected in adulthood and examined after 15 days (15dpi). p63$^+$ Krt5$^+$ tdTom$^+$ cells were found extensively in intrapulmonary airways (50.4% ± 28.3% labeling frequency) and alveoli (36.7% ± 17.6% labeling frequency) as clusters (pods) or honeycomb-like lesions (Figure 22A). By quantification, 36.7% ± 17.6% of airway Krt5$^+$ cells were found to carry the tdTom label, and 50.4% ± 28.3% of alveolar Krt5$^+$ cells were labeled, consistent with the ~50% labeling efficiency of bronchial p63$^+$ Krt5$^-$ progenitors prior to injury (Figure 22B). This lineage tracing result, combined with the adult tracing analysis, indicated that the bronchial p63$^+$ Krt5$^-$ progenitor pool responsible for H1N1-induced alveolar remodeling was established at E13.5, solely by the embryonic intrapulmonary p63$^+$ cells with no contributions from other lineages.

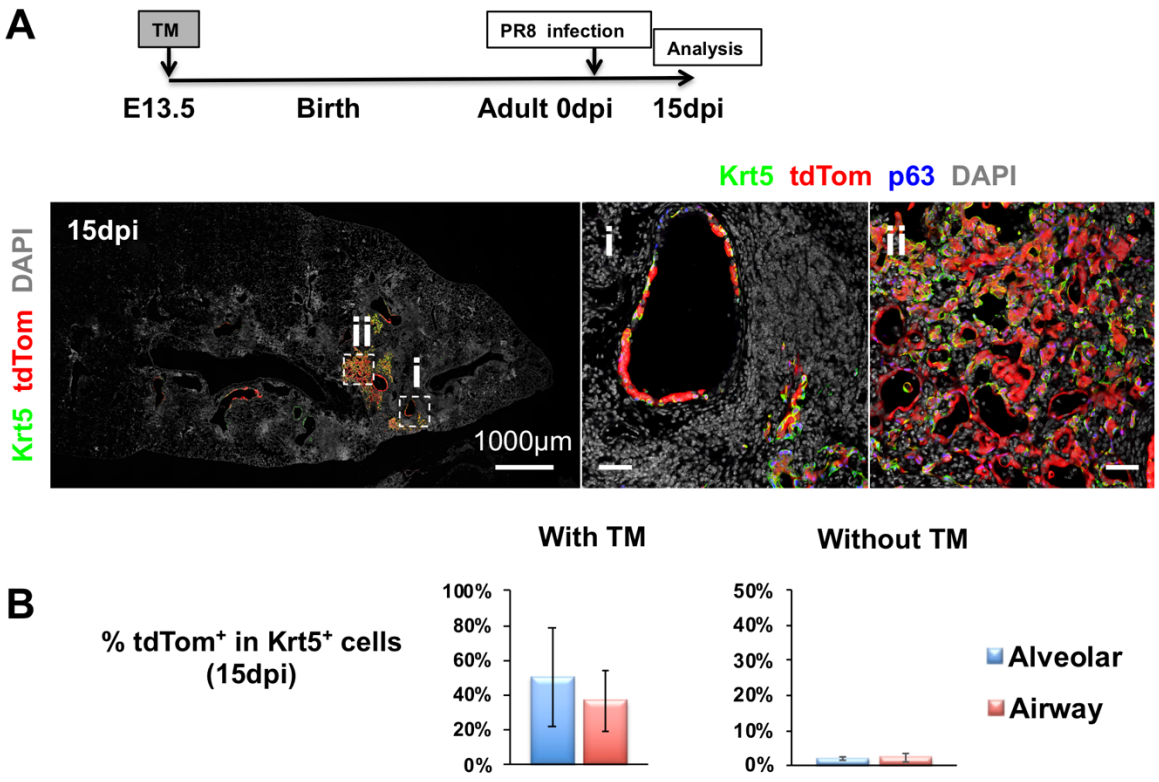


Figure 22. Intrapulmonary progenitors derived from E13.5 p63⁺ cells mediate the H1N1 aberrant alveolar remodeling.

Adult lung from mice exposed to TM at E13.5 and infected with H1N1 at 8 weeks. **(A)** IF at 15dpi: p63⁺ tdTom⁺ Krt5⁺ cells in intrapulmonary bronchi (i) and as alveolar clusters (ii. pods); boxed areas magnified in right panels. **(B)** Graphs: % tdTom⁺ Krt5⁺ cells in the H1N1-induced ectopic Krt5⁺ cells in the mice with TM treatment at E13.5 (left) or without TM exposure (right). Graphs: mean ± SEM, N⩾3. Scale bars: 50μm unless noted.

5-5. A subpopulation of p63⁺ Krt5⁻ bronchial progenitors carrying CC10 lineage label

Previous evidence of H1N1 induction of Krt5⁺ pods in CC10 lineage-labeled cells was deemed artefactual due to residual TM activity[103,105]. To circumvent this issue, adult *CC10-CreERT2; R26-tdTomato* mice were exposed to H1N1 21 days after the last TM-gavage administration. Analysis of 15dpi lungs showed robust induction of Krt5⁺ pods nearly filled the whole lung (Figure 23A). In contrast to the previous reports[105,106], extensive tdTom labeling was found in the Krt5⁺ pods. Intriguingly, more tdTom-labeled Krt5⁺ pods could

be seen in more distal regions, although still occupying mainly in the large intrapulmonary airways. Quantification showed that the tdTom$^+$ Krt5$^+$ cell clusters consisted of 30.3% ± 12.0% of the alveolar Krt5$^+$ and 45.1% ± 11.8% of the airway Krt5$^+$ cells (Figure 23B-C). Collectively, all expressed p63, indicating that CC10 lineage-labeled clusters were a subpopulation of the total H1N1-responding p63 lineage-derived cells. This suspected CC10 lineage-labeled p63$^+$ cells responsible for Krt5$^+$ pods after injury challenge were probably located more distally.

It showed that CC10 lineage-labeled cells could regenerate the alveolar compartment after H1N1 challenge[104]. To re-examine this issue, the contributions of *p63-CreERT2* and *CC10-CreERT2* lineage-labeled cells to AT1 and AT2 cells at 15dpi were compared. As mentioned above, although Pdpn$^+$ squamous cells could be identified with tdTom in *p63-CreERT2* tracing, these cells were unlikely to be newly generated AT1 cells. Rarely, one pro-Spc$^+$ tdTom$^+$ Pdpn$^-$ AT2 cell (pointed by white arrow in Figure 23E left bottom panel) and one pro-Spc$^+$ tdTom$^+$ Pdpn$^+$ cell (pointed by magenta arrowhead in Figure 23E left bottom panel) were identified, probably transitioning from a Krt5$^+$ pod cell to an AT2 cell. In contrast, in *CC10-CreERT2* tracing, more Pdpn$^+$ tdTom$^+$ cells could be identified. These cells were probably AT1 cells as they had long and thin extension, with typical AT1 cellular morphology (pointed by white arrow in Figure 23E right upper panel). Additionally, more pro-Spc$^+$ Pdpn$^-$ AT2 cells could be found bearing the lineage label, suggesting a significant contribution from CC10 lineage to alveolar regeneration. Nevertheless, *CC10-CreERT2* labeled a big proportion of endogenous AT2 cells (Figure 23E right bottom panel). A recent paper showed evidence that a Wnt responsive subpopulation of AT2 cells (Axin2$^+$ alveolar epithelial progenitors) plays major role in generating the nascent AT2 and AT1 cells during H1N1 injury repair[89]. It cannot be distinguished whether the contribution from CC10 lineage to alveolar repair was from the airway secretory cells or CC10 lineage-labeled AT2 cells.

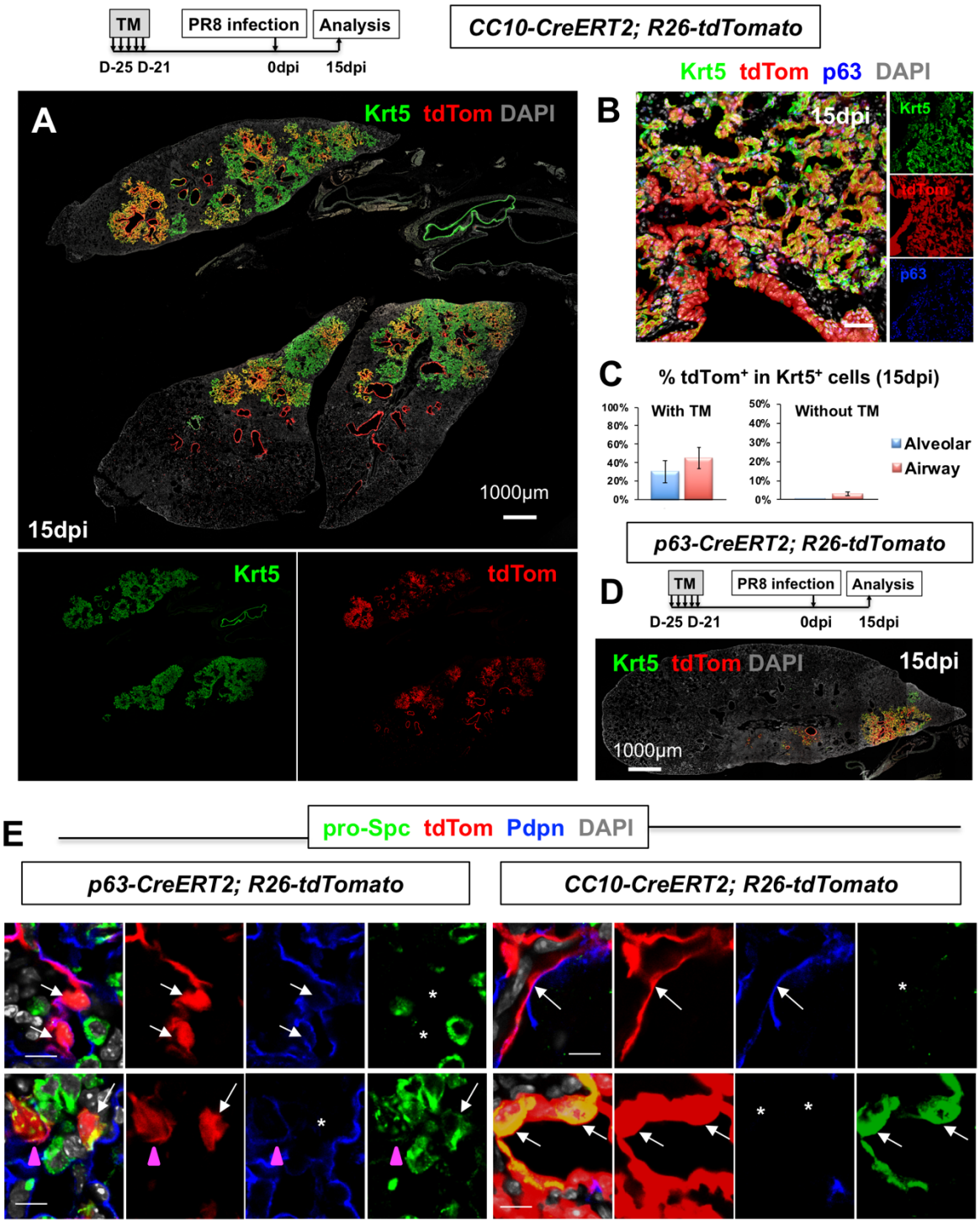


Figure 23. Intrapulmonary p63^{+} Krt5^{-} progenitors include a subpopulation of CC10 lineage-labeled cells responsible for H1N1 induction of pods.

(A) *CC10-CreERT2* adult mice exposed to H1N1 after 21days of TM exposure. IF: lung sections at 15dpi with CC10 lineage-labeled airways and $Krt5^+$ clusters overlapping partially with lineage-negative $Krt5^+$ clusters (green). **(B)** IF: $tdTom^+$ "trail" expanding to alveolar compartment. **(C)** Graphs: % $tdTom^+$ $Krt5^+$ cells in the H1N1 induced ectopic $Krt5^+$ cells in the mice with TM treatment in adulthood (left) or without TM exposure (right). **(D)** *p63-CreERT2* adult lung at 15dpi: IF: less pods compared to *CC10-CreERT2* suggestive of a more attenuated response to H1N1. **(E)** IF analysis of 15 dpi lungs showing rare tdTom-labeled alveolar Type I ($Pdpn^+$ pro-Spc^- with line structure, upper panels) and Type II cells ($Pdpn^-$ pro-Spc^+, lower panels) in both *p63-CreERT2* and *CC10-CreERT2* mice. *no fluorescent signal. Magenta arrowhead showing the $tdTom^+$ intermediate cell co-expressing Pdpn and pro-Spc. Scale bars: 50μm in **(B)**, 10μm in **(E)** unless noted.

The possibility was tested that CC10 lineage-labeled $p63^+$ $Krt5^-$ cells were already present in the adult lung prior to H1N1 infection. Remarkably, IF of TM-treated uninjured adult *CC10-CreERT2* mice revealed about 31.9% ± 4.3% of the $p63^+$ $Krt5^-$ cells scattered in the main intrapulmonary bronchi labeled by tdTom (Figure 24A-B). As ~30% of bronchial $p63^+$ $Krt5^-$ progenitors were labeled in *CC10-CreERT2* mice prior to injury, and ~30% of ectopic $Krt5^+$ cells carried the lineage label at 15dpi, it could be inferred that the CC10 lineage-labeled subpopulation of intrapulmonary $p63^+$ $Krt5^-$ progenitors were responsible for the lineage-labeled $Krt5^+$ pods and required to mount a full response to H1N1.

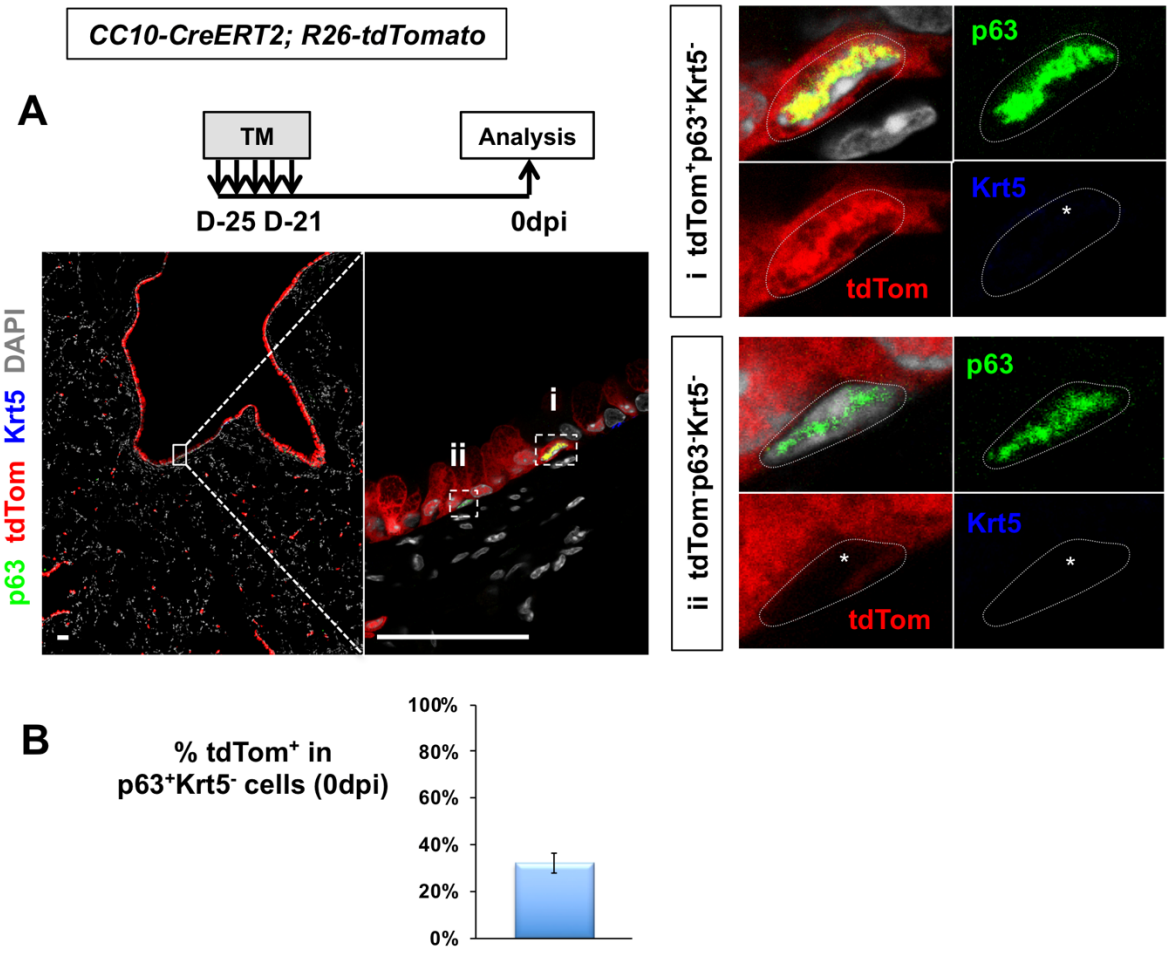


Figure 24. A subpopulation of intrapulmonary p63⁺ Krt5⁻ progenitors could be labeled by *CC10-CreERT2* lineage.

(**A**) Uninjured adult *CC10-CreERT2* mice 21 days after Tm: rare CC10 lineage-labeled p63⁺ Krt5⁻ progenitors in intrapulmonary main bronchi (i) near unlabeled p63⁺ Krt5⁻ cells (ii). **(B)** Graph: % tdTom⁺ p63⁺ Krt5⁻ cells in total intrapulmonary p63⁺ Krt5⁻ cells. Graph: mean ± SEM, N⩾3. Scale bars: 50μm unless noted.

Moreover, the labeling pattern of H1N1-exposed *p63-CreERT2* and *CC10-CreERT2* differed dramatically in the trachea. Each reporter labeled its respective cell type in TM-treated uninjured trachea (Figure 25A-C). By contrast, after injury *p63-CreERT2* mice labeled almost all the epithelial cells in both basal and luminal compartments but none in *CC10-CreERT2* mice (Figure 25B-C). This confirmed the stem cell

identity of the adult tracheal p63^{+} Krt5^{+} tdTom^{+} in post-injury repair and further supported the idea that the progenitor cells sharing both CC10 and p63 lineages reside in intrapulmonary airways, not the trachea.

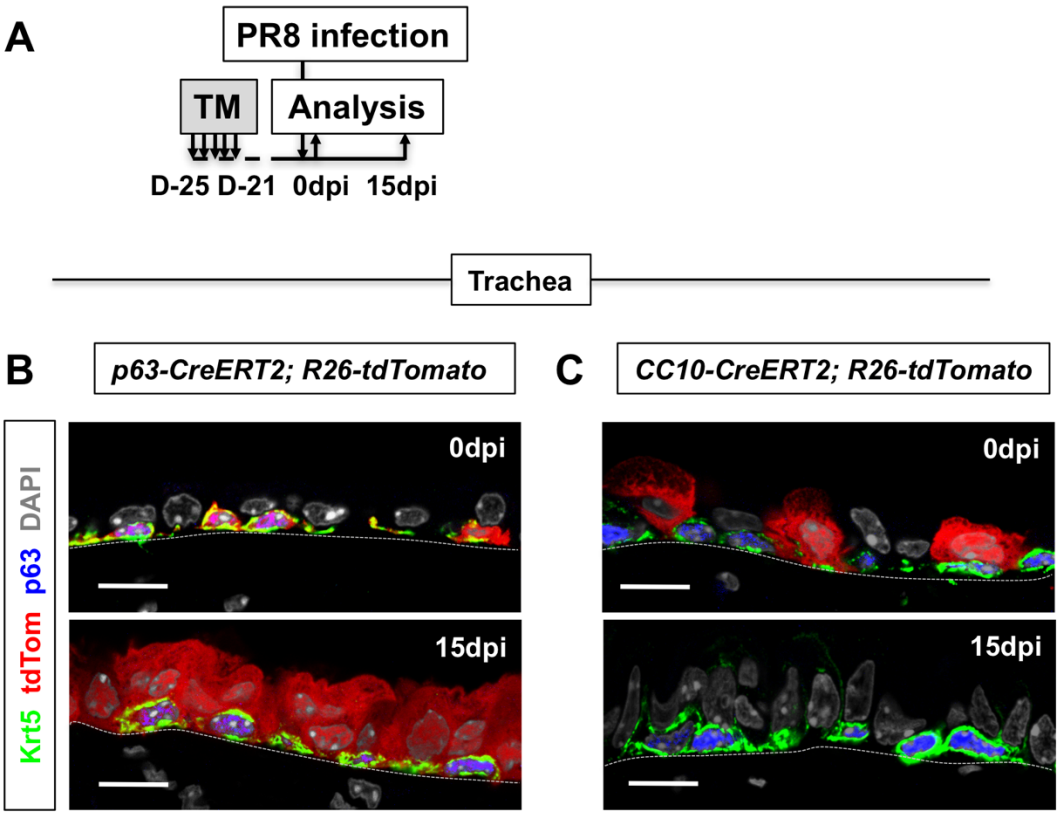


Figure 25. Spatial restriction of CC10 lineage-labeled p63^{+} cells inside the lung.

(**A)** Adult *p63-CreERT2; R26-tdTomato* or *CC10-CreERT2; R26-tdTomato* mice were exposed to TM by oral gavage and then subjected to H1N1 challenge 21 days after the last dose of TM administration. **(B-C)** Tracheal sections of adult *p63-CreERT2* (left panel) or *CC10-CreERT2* (right panel) mice showing no p63^{+} CC10 lineage-labeled cells in the uninjured tracheal epithelium (present in the uninjured intrapulmonary epithelium). Bottom panels: H1N1 injury results in extensive epithelial repair at 15dpi by p63 lineage-labeled but not CC10 lineage-labeled tdTom^{+} cells in trachea. Scale bars: 10μm.

The extensive labeling pattern in the H1N1-induced ectopic Krt5^{+} cells was not observed in other Cre lines (Spc, Upk3a, Notch3-driven) using the same TM regimen (Figures 26A-C), further suggesting the overlap between CC10 and p63 lineages were not artificial products in lineage tracing experiments.

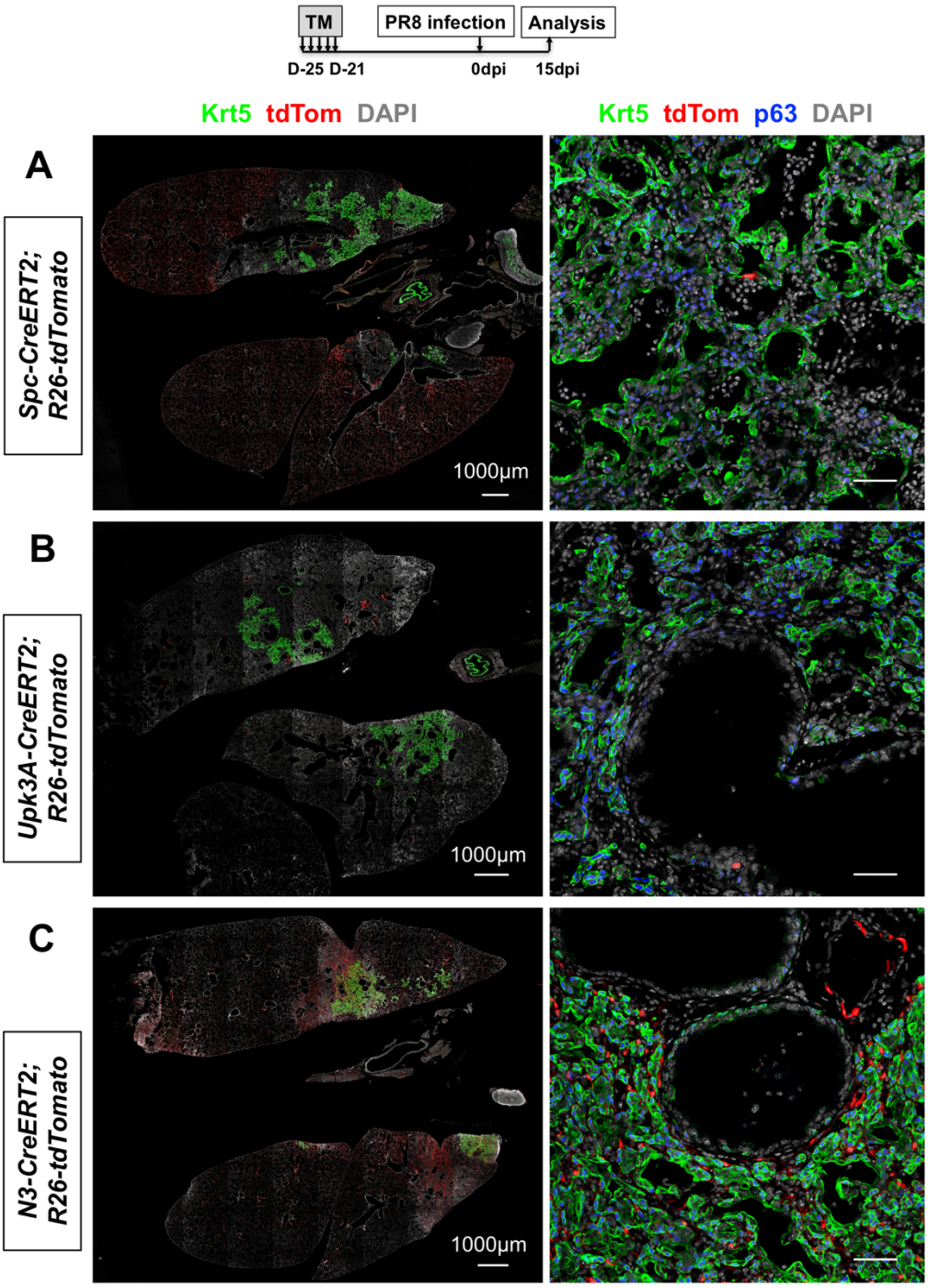


Figure 26. Variable Cre lines of lung lineages show minimal labeling in the H1N1-induced ectopic Krt5$^+$ cells.

(A-C) IF analysis of 15dpi lungs showing non-overlapped pattern between tdTom lineage label and Krt5 cells. Left panel showing tile scanning images of whole lung sections; right panels showing areas of Krt5$^+$ pods. **(A)** *Spc-CreERT2* line labeling the descendants of alveolar Type II cells. N=2 from two independent experiments. **(B)** *Upk3A-CreERT2* line labeling the descendants of secretory cells associated with neuroendocrine bodies and terminal bronchioles. N=5 from 3 independent experiments. **(C)** *N3-CreERT2* line labeling the descendants of mesenchymal and scattered epithelial cells in airways. N=3 from one experiment. Scale bars: 50μm unless noted.

5-6. p63 haploinsufficiency in H1N1 injury response

A comparison of H1N1-infected *CC10-CreERT2* and *p63-CreERT2* mice suggested a more attenuated response of the *p63-CreERT2* knock-in mice with less pods (Figure 23A & D). Therefore, p63 haploinsufficiency was hypothesized to influence this response. Notably, the H1N1-infected CC10 reporter mice showed Krt5$^+$ pods at a significantly higher frequency (100%), some slightly less proximal compared to *p63-CreERT2* mice at similar dpi. Moreover, in average 20 p63$^+$ Krt5$^-$ cells per lung section could be identified from the adult *CC10-CreERT2* mice during homeostasis, while this number decreased to 3 in *p63-CreERT2* mice, suggesting an allelic function of p63 in generating or maintaining this progenitor pool in intrapulmonary airways.

To avoid confounding factor from mouse background, the injury response between *p63-CreERT2; R26-tdTomato* mice were compared with their *p63+/+; R26-tdTomato* littermates. Similar to the *CC10-CreERT2* mice which possessed two functional alleles of p63 gene, the *p63+/+; R26-tdTomato* littermates produced massive Krt5$^+$ pods in the lung, while *the p63-CreERT2; R26-tdTomato* mice showed much fewer pods with more enrichment in the proximal part of lung (Figure 27A &C). This difference was also found in other CreERT2 knock-in lines tested in which disruption of the endogenous gene did not affect the progenitor pool (Figure 26A-C). About 60% of the p63 heterozygous mice failed to generate ectopic Krt5$^+$ cells after injury, although their lung experienced similar levels of injury shown by collapse of alveolar structure and massive infiltration of immune cells (Figure 27B). In those *p63-CreERT2* mice showing Krt5$^+$ pods, all

ectopic Krt5$^+$ cells were Sox2$^+$ and Nkx2.1$^+$, and these pods were closely associated with the regions of most severe alveolar damage as shown by the lack of Sox2$^+$ or Nkx2.1$^+$ lung epithelial cells (Figure 27A).

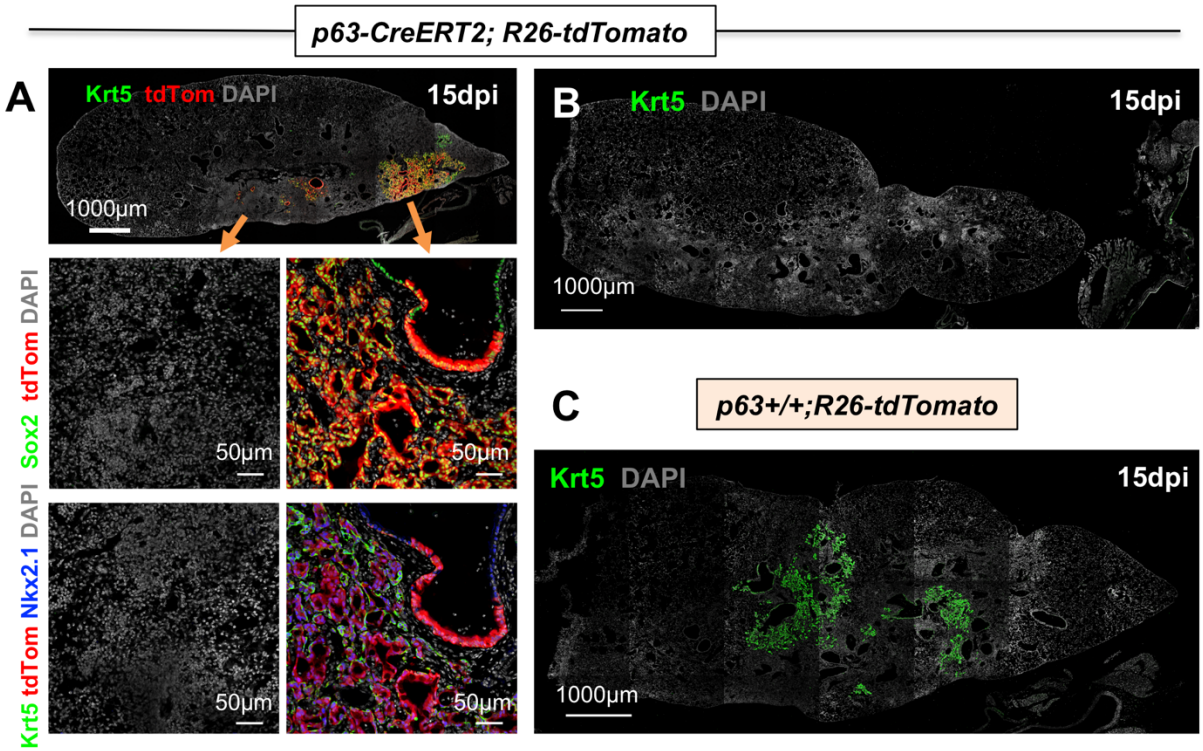


Figure 27. Dramatically attenuated response to H1N1 challenge due to p63 haploinsufficiency.

(A) IF analysis of 15dpi lung sections of *p63-CreERT2* mice: cells in the pods are Sox2$^+$ and Nkx2.1$^+$, while the surrounding parenchyma devoid of pods is absent of lung epithelial cells. **(B-C)** IF analysis of 15dpi lung sections of *p63-CreERT2* mice or the *p63+/+* WT littermates. N=10 for *p63+/+* animals showing ectopic Krt5$^+$ cells in lung after H1N1 challenge in two independent experiments. Note some *p63-CreERT2* mice failed to generate the ectopic Krt5$^+$ cells in response to H1N1 challenge, while this phenomenon was never observed in *p63+/+* WT animals.

5-7. p63 lineage contribution to a new "variant" neuroendocrine cells

A collaboration study with Dr. Xingbin Ai's group revealed that pulmonary neuroendocrine cells are heterogeneous in expression of CGRP and Tubb3. Tubb3 is a neuron-specific tubulin beta class III marker.

The neuroendocrine cells, especially those in NEBs, in mouse lungs are CGRP$^+$, while in trachea, extremely rare neuroendocrine cells could be identified near the submucosal glands, expressing Tubb3. Some of these Tubb3$^+$ cells also expressed CGRP at low levels.

Cell culture experiments suggested that airway BCs could generate neuroendocrine cells (CGRP$^+$ Tubb3$^+$ or CGRP$^-$ Tubb3$^+$). To investigate whether airway BCs served as the progenitors of tracheal neuroendocrine cells *in vivo*, the lineage labeling pattern was analyzed in the *p63-CreERT2; R26-tdTomato* mice with TM exposure at E17.5. As described before, E17.5 tracing labeled specifically p63$^+$ Krt5$^+$ Krt8$^-$ preBC population at E18.5 with a high efficiency of 95.6% (Figure 3). By P21, extensive tdTomato-labeled CGRP$^+$ Tubb3$^+$ cells (marked by white arrow in Figure 28B) and CGRP$^-$ Tubb3$^+$ cells (marked by asterisk in Figure 28B) could be identified. These suggested that the rare neuroendocrine cells in trachea were derived from preBCs at E18.5 and were maintained by airway BCs during homeostasis.

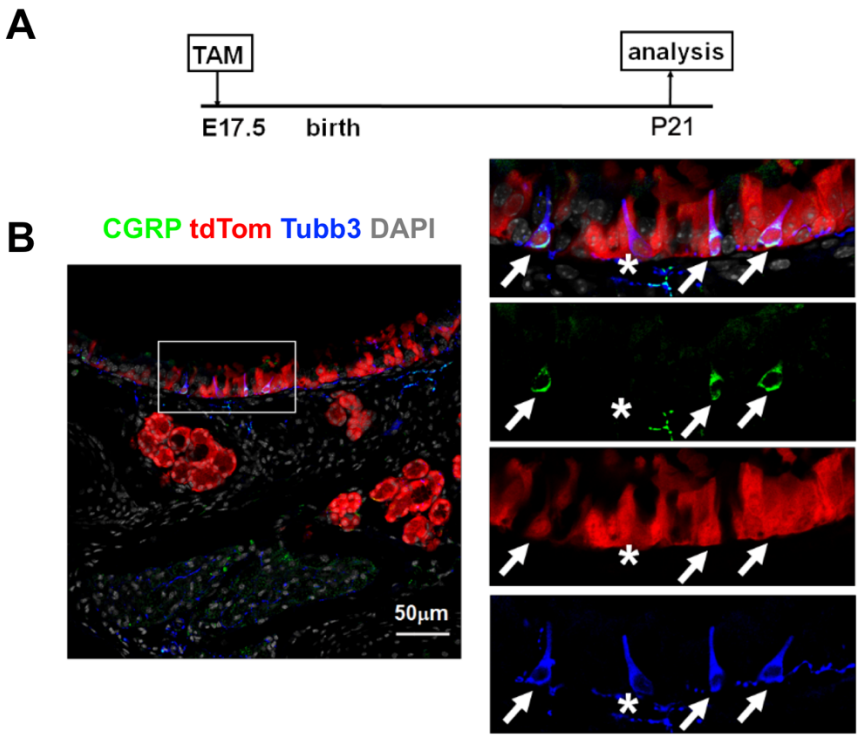


Figure 28. Neuroendocrine cells in trachea were derived from E18.5 p63$^+$ preBCs.

(A) *p63-CreERT2; R26-tdTomato* mice were exposed to TM at E17.5 and analyzed for lineage labeling at P21. **(B)** IF of tracheal sections showing tdTom-labeled CGRP$^+$ Tubb3$^+$ neuroendocrine cells (arrows) and CGRP$^-$ Tubb3$^+$ neuroendocrine cells (asterisks).

To investigate whether the intrapulmonary p63$^+$ Krt5$^-$ progenitors contribute to the canonical neuroendocrine cells during homeostasis and injury response, *p63-CreERT2; R26-tdTomato* mice were exposed with TM, and the tdTomato labeling was analyzed at 0dpi (21 days after the last dose of TM to ensure TM clearance) and 15dpi (Figure 29A). Canonical neuroendocrine cells in intrapulmonary airways did not express Tubb3 either prior to or after injury challenge (Figure 29B-C). There was minimal evidence supporting the contribution from p63 lineage to neuroendocrine cells at either 0dpi or 15dpi (Figure 29B-C).

E13.5 TM exposure showed again no tdTomato$^+$ CGRP$^+$ cells (Figures 29E-F), although massive labeling could be observed in the surrounding Krt5$^+$ pods. Given the low recombination efficiency of intrapulmonary p63$^+$ progenitors at the embryonic stage, and the rare frequency of pulmonary neuroendocrine cells, p63 lineage probably had minimal to none contribution to the canonical neuroendocrine lineage inside the lung.

To investigate whether secretory lineage contributes to the canonical neuroendocrine cells, lineage tracing was performed using *CC10-CreERT2; R26-tdTomato* mice with TM administration in adulthood (Figure 29G). As discussed above, the labeling efficiency was high enough to label the variant secretory cells surrounding NEBs (they express much low levels of CC10 comparing to the canonical secretory cells) (Figure 29H). Again CGRP$^+$ Tubb3$^-$ canonical neuroendocrine cells were not labeled in this system either prior to or after injury challenge (Figures 29H-I). Furthermore, in the time course analysis after injury challenge, co-expression of Krt5 and CGRP was never observed at all stages analyzed. A significant amount of ectopic Krt5$^+$ cells lining the intrapulmonary airways proliferated actively before 15dpi, shown by Ki67 staining (Figures 30A-D). But only extremely rare Ki67$^+$ CGRP$^+$ cells could be identified. Therefore, neither p63 lineage nor CC10 lineage contributed to the canonical CGRP$^+$ Tubb3$^-$ neuroendocrine population in mouse intrapulmonary airways.

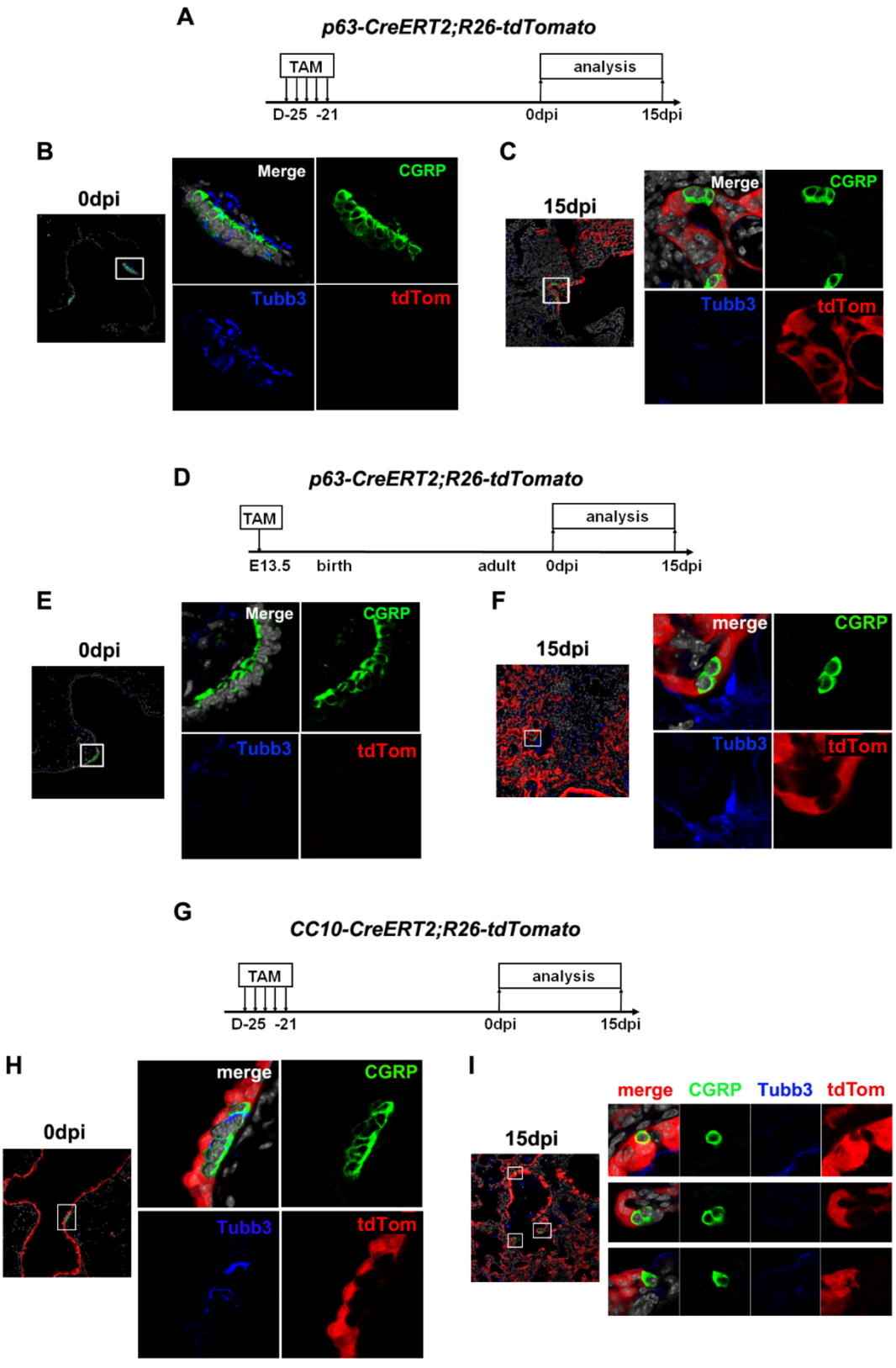


Figure 29. p63 lineage and CC10 lineage could not label classic PNECs in mouse lung.

(A, D, G) Lineage tracing strategies in adulthood or embryonic stage**. (B, C, E, F, H, I)** IF analysis of 15dpi lung sections showing nearly no tdTom labeling in the CGRP^{+} PNECs at 0dpi or 15dpi.

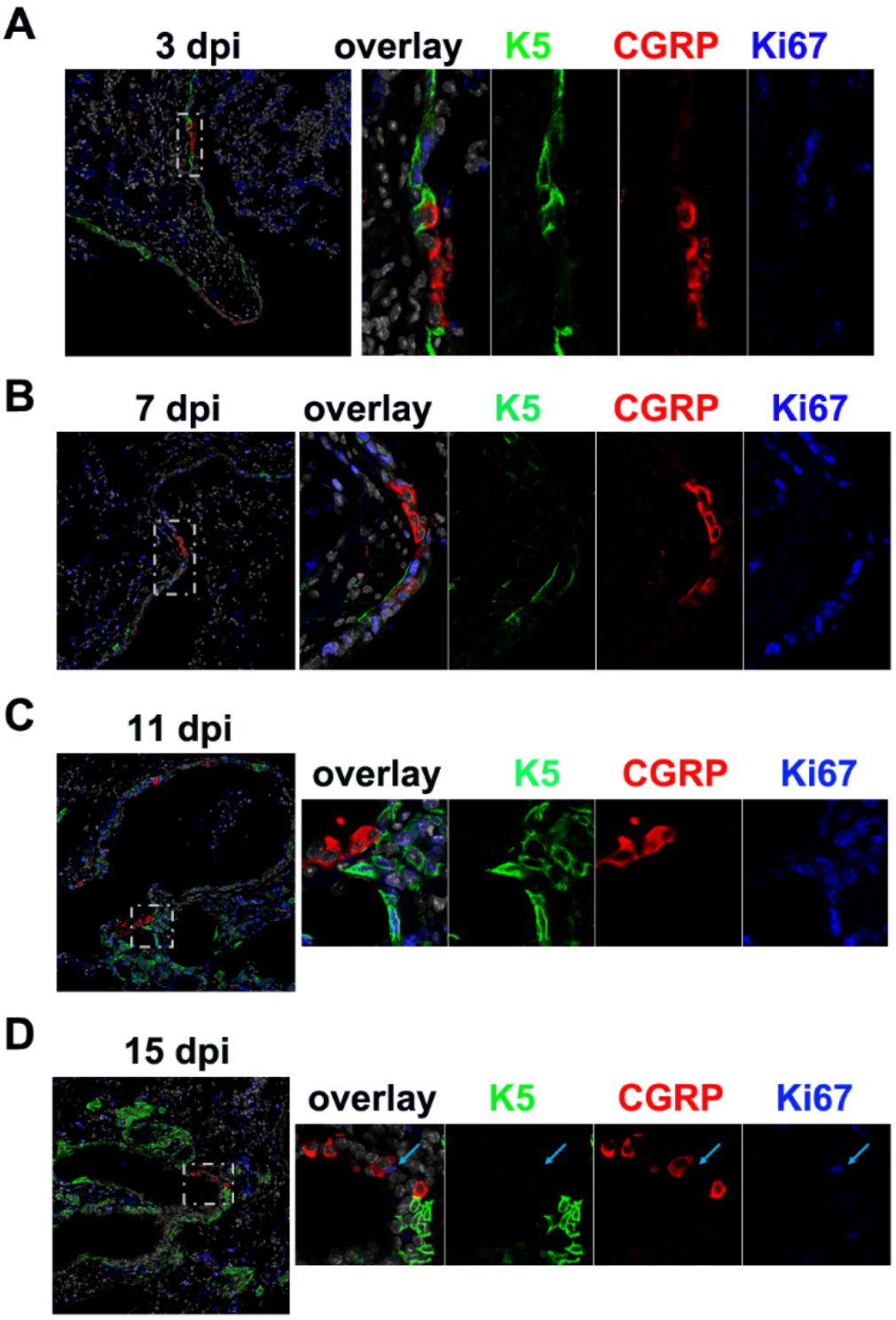


Figure 30. Little proliferation of PNECs after H1N1 challenge.

(A-D) IF analysis of WT mouse lungs harvested at different time points following H1N1 infection. Boxed areas were enlarged on the right panels with separated channels. Cyan arrow in **(D)** showing a $Ki67^{+}$ $CGRP^{-}$ cells.

Chapter 6. Molecular characterization of p63^{+} progenitors during development and homeostasis

6-1. Transcriptome analysis of airway BCs during postnatal maturation process.

The lineage tracing analysis performed in embryonic *p63CreERT2; R26-tdTomato* animals indicated that in trachea preBCs at E18.5 self-renew to generate the adult basal stem cell pool. These preBCs at E18.5, although sharing multiple markers with adult airway BCs like p63, Krt5, Itgb4 (CD104), gradually acquire additional BC markers (e.g. NGFR) during postnatal development. This suggests a maturation process from preBCs to adult BCs. In addition, E18.5 preBCs have a higher proliferative rate (8% at E18.5 vs. 2% in adulthood; Figures 7A & 36B) and contribute more rapidly to the luminal lineages (Figures 3 & 25B), to elongate the growing tracheal tube during postnatal development. Aiming at revealing the molecular mechanisms regulating this BC maturation process, transcriptome analysis was performed to generate the gene expression profiles of E18.5 preBCs and adult TrBCs. Comparison between the gene expression patterns of these two populations may shed light on how adult stem cells develop and become quiescent.

To obtain pure cell populations for E18.5 preBCs, *p63-CreERT2; R26-tdTomato* mice were exposed to TM at gestation day 17.5 and embryonic tracheas were isolated one day later. Adult TrBCs were isolated with TM treatment in adulthood and a chase time of 3 days. Both BC populations were isolated and sorted using both tdTomato and surface markers (EpCAM^{+}; Lin^{-}; tdTomato^{+}; CD104^{hi}) (Figure 31A). Cytospin results of sorted populations showed more than 90% of cells as p63^{+}, and more than 95% as Krt5^{+} (transcription factors such as p63 are sensitive for sample preparation while intermediate filaments such as Krt5 are more stable) (Figure 31B).

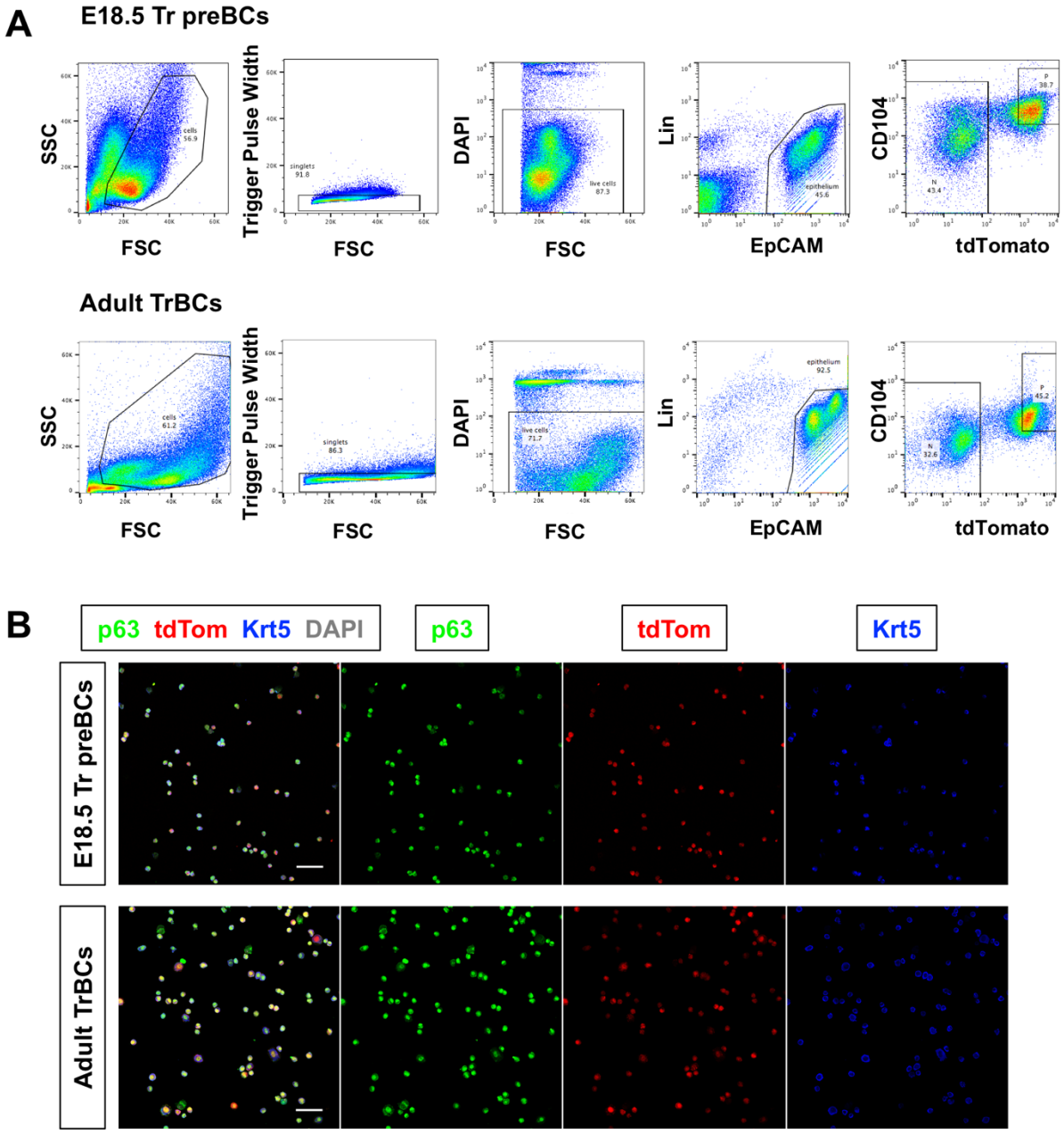


Figure 31. Isolation of E18.5 preBCs and adult BCs from tracheal epithelium.

(A) Gating strategy to sort Lin^{-} EpCAM^{+} tdTom^{+} CD104^{hi} populations. **(B)** IF analysis of sorted cells after cytospin showing high purity based on p63 and Krt5 expression. Scale bars: 50μm.

E18.5 preBCs exhibited a distinct gene expression profile from adult TrBCs. For both populations, more than 83% of the sequencing reads were pseudoaligned to the genome (Figure 32A). Principal component analysis (PCA) was performed on all 12 transcriptomes, and this clustered all E18.5 preBCs together in contrast to other adult TrBC samples (Figure 32B). 658 genes were significantly enriched in E18.5 preBCs ($q<0.01$; $b>1.5$), while 524 genes were selectively enriched in adult TrBCs ($q<0.01$; $b<-1.5$) (Figure 32C). Gene ontology analysis revealed E18.5 preBCs overrepresented gene sets involved in “Extracellular structure organization”, “Tissue development”, “Organ morphogenesis”, and so on, while adult TrBCs showed overrepresented gene sets involved in “Immune system process”, “Defense response”, “Response to external stimulus”. Notably, gene sets for “Regulation of cell differentiation” and “Regulation of cell proliferation” were highly enriched in E18.5 preBCs, consistent with their higher capacities of self-renewal and differentiation (Figure 32D). By contrast, adult TrBCs in homeostasis contained subpopulations of cells more quiescent or transitioning towards luminal commitment, expressing transcripts shared with secretory or multiciliated cells.

Gene set overlapping analysis of KEGG pathways revealed that components of Hedgehog, WNT, TGFβ and insulin growth factor (IGF) signaling pathways were significantly enriched in E18.5 preBCs, while adult TrBCs were enriched in Jak/Stat pathway (Figures 32D, 32F). Shh and Gli2 showed considerable expression levels in E18.5 preBCs (Figure 32F). As a downstream target of both Shh and TGFβ pathways[123], Gli2 was reported to promote cell proliferation and to inhibit apoptosis, and to play important roles in multiple squamous cell cancers including lung squamous cell carcinoma and skin basal cell carcinoma[124,125]. Interestingly, hypoxia has been suggested to induce the transcription of Gli2 through Hif1α[126]. Cells at prenatal stages are generally under hypoxic environment. This might underlie the intrinsic difference between E18.5 preBCs and adult TrBCs in proliferative activities. Additionally, several Wnt ligands (Wnt5b, Wnt11, Wnt10b, Wnt6) were over-represented in E18.5 preBCs. The expression of Axin2 and Lef1 suggested activation of canonical Wnt signaling pathway in these embryonic cells. Insulin signaling pathway was also activated, consistent with the fact that preBCs are responsible for populating both basal and luminal compartments during postnatal growth. Multiple BMP/TGFβ antagonists, Chrd, Fstl1, Fst, were enriched in E18.5 preBCs, suggesting suppression of BMP/TGFβ in these cells. Inhibition of BMP and TGFβ in adult epithelial BC culture enhances self-renewing capacity and represses spontaneous

differentiation[62], and BMP signaling constrains cellular proliferation in adult BCs in homeostasis[127]. Here the data suggested that embryonic preBCs may employ the same strategy to maintain the progenitor status. On the other hand, high activity of Jak/Stat signaling pathway in adult TrBCs reflect that these adult cells are more prone to commit to differentiation and less to self-renew[55]. Interestingly, Socs3, as a negative regulator for Jak/Stat pathway[128], was enriched in E18.5 preBCs, consistent with the difference observed in the transcriptomes of TrBCs and their embryonic precursors.

Examination of the most differentially expressed genes confirmed previous report that Ngfr is only expressed postnatally (Figure 32E). Ndufa4l2, a gene highly induced by HIF-1α[129], is the most significantly enriched gene in E18.5 preBCs. This might reflect the major difference of the two cells in environment oxygen before and after birth. Transcription factors found in E18.5 preBCs included Sox11, Hif3a, Mycn, Foxa2. The ETS-related transcription factors, Etv4 and Etv5, described as squamous cell oncogenes[130,131], were highly enriched in E18.5 preBCs (Figure 32G).

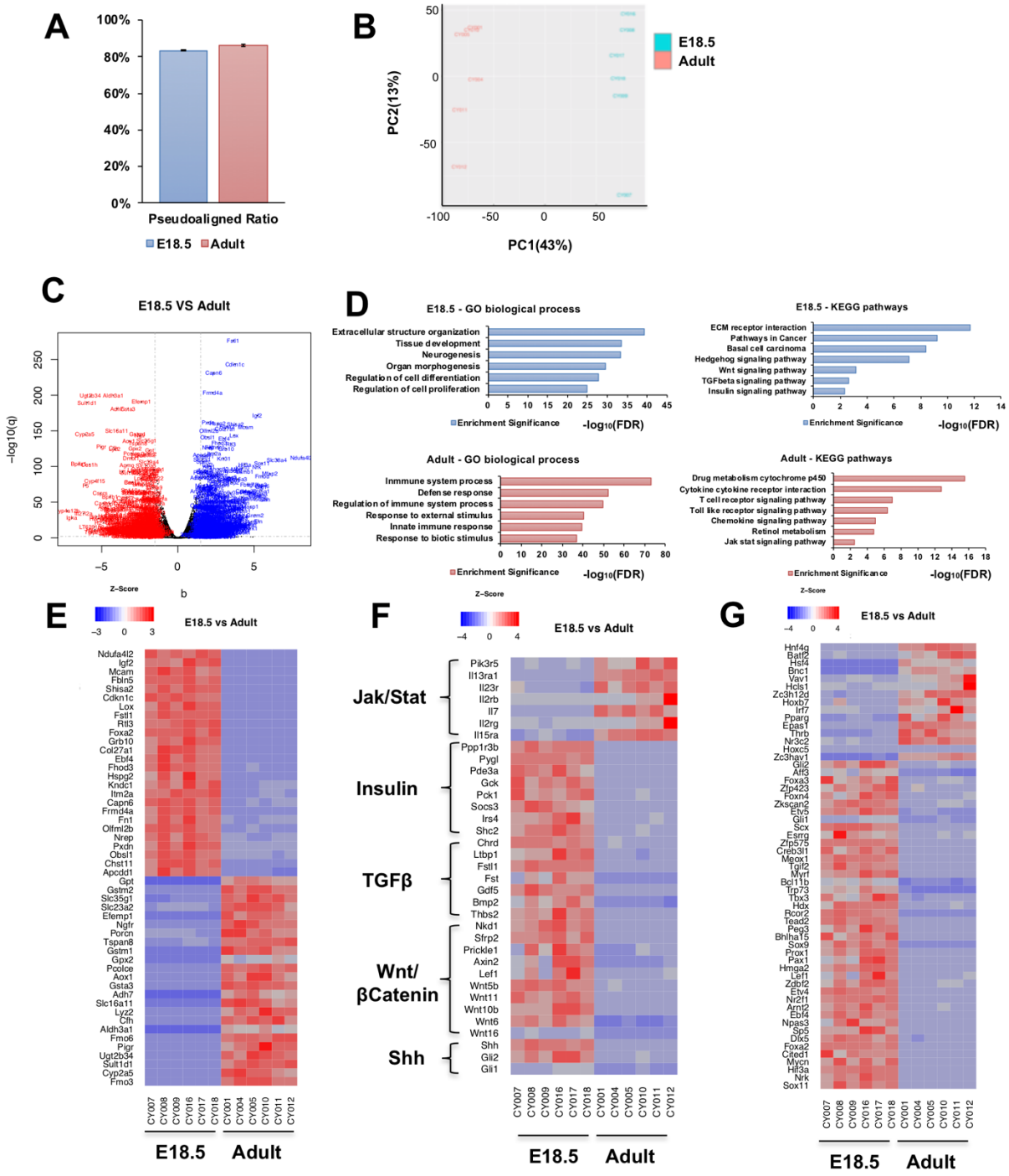


Figure 32. Transcriptome profiles of E18.5 preBCs and adult TrBCs.

(A) Bar graph showing the pseudoaligned ratios of E18.5 preBCs and adult TrBCs in an RNA-seq quantification program, Kallisto. **(B)** Principal component analysis (PCA) showing distinct difference between E18.5 and adult samples. **(C**) Volcano plot showing differentially expressed genes in E18.5 preBCs and adult TrBCs represented in black dots; genes significantly enriched in E18.5 preBCs were labeled with gene names in blue ($-\log_{10}(q)>2$; $b>1.5$); genes significantly enriched in adult TrBCs were labeled with gene names in red ($-\log_{10}(q)>2$; $b<-1.5$). **(D)** Enriched pathways in each population. Representative MSigDB gene sets (GO biological process & KEGG pathways) in E18.5 preBCs (blue) and adult TrBCs (red). **(E)** Top 50 genes ranked by significance showing dramatic difference in expression pattern in E18.5 preBCs and adult TrBCs. **(F)** Distinct expression pattern of genes in each representative KEGG pathways. **(G)** E18.5 preBC-specific and adult TrBC-specific transcription factors. **(E-G)** Relative expression levels of genes (row-wise Z-score of mean tpm). N=6 for each population.

6-2. Transcriptome analysis of adult tracheal BCs and esophageal BCs in homeostasis

$p63^+$ $Krt5^+$ BCs are adult stem cells responsible for homeostatic maintenance and injury repair in multiple epithelial systems. The stem cell features of self-renewal and differentiation can be hijacked during disease or injury-repair. To identify the intrinsic regulators of the programs of tissue-specific lineage commitment, transcriptome analysis of tracheal BCs and esophageal BCs was performed in adult mice in homeostasis. The focus on these two BC types was based on four reasons: 1. Although the integrity of both epithelia is maintained by BCs, the tracheal epithelium differs from the esophageal epithelium in turn-over rate, epithelial architecture and cellular composition of the luminal compartment. Tracheal BCs proliferate slower and generate secretory and multiciliated cells that populate the pseudostratified epithelium. In contrast, the esophageal BCs proliferate much faster, and give rise to a squamous multilayered stratified epithelium. 2. Trachea and esophagus share a common developmental origin from the anterior foregut endoderm, which undergoes dorsal-ventral patterning to specify the ventral trachea and dorsal esophagus. 3. Trachea and esophagus develop closely associated with each other, and are likely to have been exposed common signals, which might underlie a reciprocal cellular plasticity between these two BC populations. 4. A number of studies show that the respiratory epithelium can undergo metaplastic changes and convert into a squamous stratified epithelium.

Pure TrBCs and esoBCs were obtained using the same sorting strategy as described above based on Lin- EpCAM^{+} tdTomato^{+} CD104^{hi} (Figure 33A) and confirmed by cytospin (Figure 33B). For both BC populations, more than 86% of the sequencing reads were pseudoaligned to the genome (Figure 34A). TrBCs and esoBCs, although sharing many canonical BC markers and adult stem cell features, exhibited significant differences in their transcriptomes, analyzed by PCA (Figure 34B). 1081 genes were selectively over-expressed in TrBCs (q<E-10; b>2), and 661 genes were highly expressed in esoBCs (q<E-10; b<-1) (Figure 34C). Gene ontology analysis showed that TrBCs over-represented gene sets included “Immune system process”, “Response to external stimulus”, “Defense response”. EsoBCs showed overrepresentation in genes related to “Epithelial development”, “Regulation of cell proliferation”, “Epidermis development”, “Ribosome biogenesis”. This was consistent with the known difference in turn-over rates between tracheal and esophageal endoderm (Figure 34D). Gene set overlapping analysis of KEGG pathways revealed that components of Hedgehog signaling pathway, basal cell carcinoma (BCC), and pathways in cancer were significantly enriched in esoBCs, while TrBCs have a transcriptome enriched in genes annotated with activated MAPK signaling pathway (Figures 34D & F).

The leucine-rich repeat-containing G-protein coupled receptors, Lgr5 and Lgr6, are the receptors for R-spondins, which potentiates the WNT signaling. Intriguingly, Lgr6 is among the most differentially expressed genes enriched in TrBCs, even though multiple WNT signaling pathway components (Wnt3a, Wnt10a, Wnt16, Fzd10) were highly expressed in esoBCs but not TrBCs (Figures 34E & F). Both Lgr5 and Lgr6 have been described as stem cell markers in several different epithelial systems such as intestine, epidermis, and sebaceous gland. In the respiratory system, the mesenchymal populations expressing Lgr5 and Lgr6 have been reported to promote alveolar differentiation and airway differentiation, respectively, through mesenchymal-epithelial crosstalk[132]. However, the roles of Lgr5 and Lgr6 in epithelial compartment remained unexplored. In our transcriptome analysis, we found that Lgr5 was specifically expressed in E18.5 tracheal preBCs, although at low levels. Neither adult TrBCs nor esoBCs expressed Lgr5, indicating a temporal expression pattern of Lgr5 in TrBCs. Lgr6 was absent from adult esoBCs, but found at high levels in tracheal E18.5 preBCs and adult TrBCs, suggesting a tissue-specific expression pattern (Figure 34E). Further characterization of Lgr5 and Lgr6 expressing epithelial populations may reveal the heterogeneity in BC populations regarding to stemness or differentiation potentials.

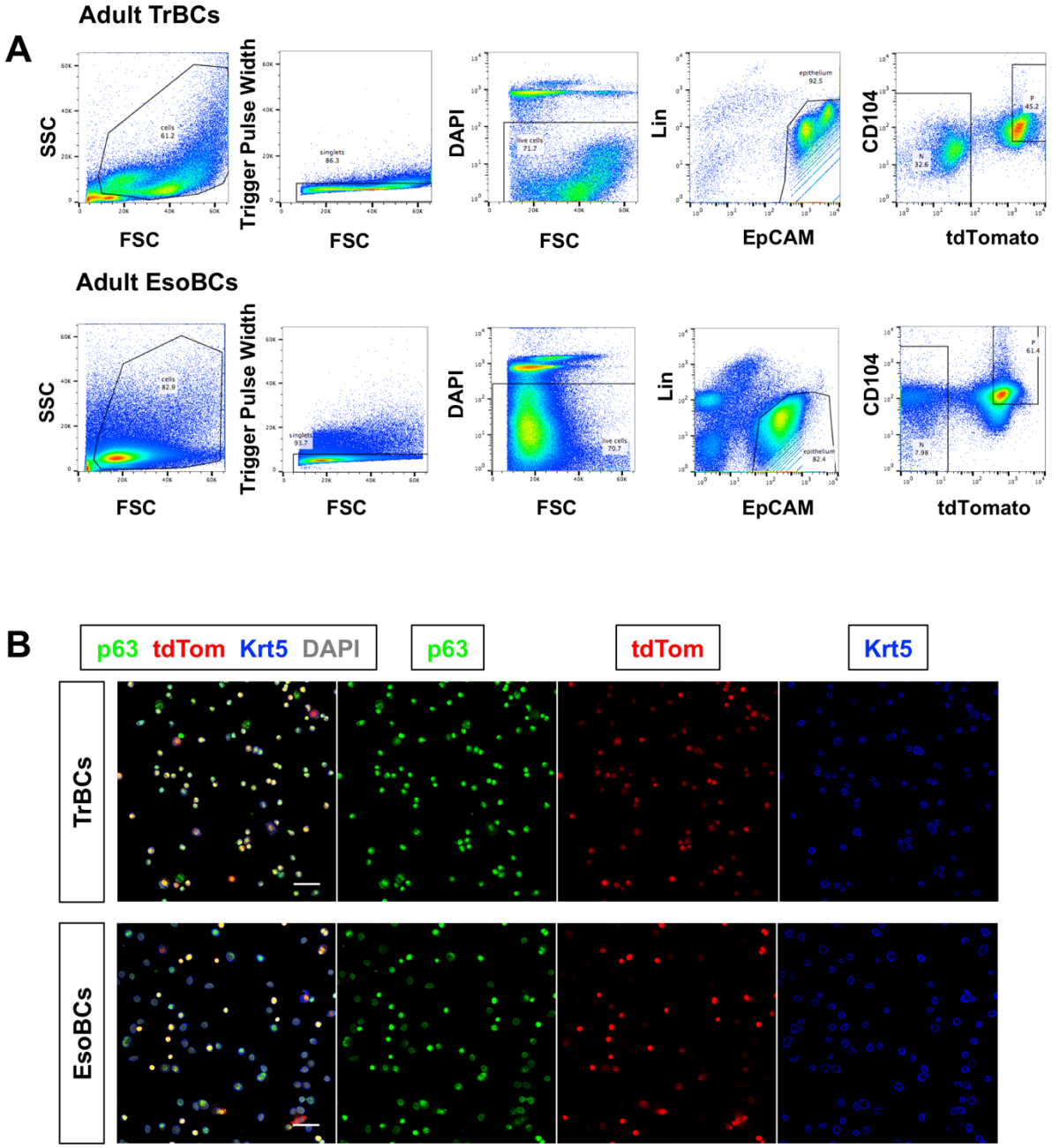


Figure 33. Isolation of adult BCs from tracheal and esophageal epithelia.

(A) Gating strategy to sort Lin^{-} EpCAM^{+} tdTom^{+} CD104^{hi} populations. **(B)** IF analysis of sorted cells after cytospin showing high purity based on p63 and Krt5 expression. Scale bars: 50μm.

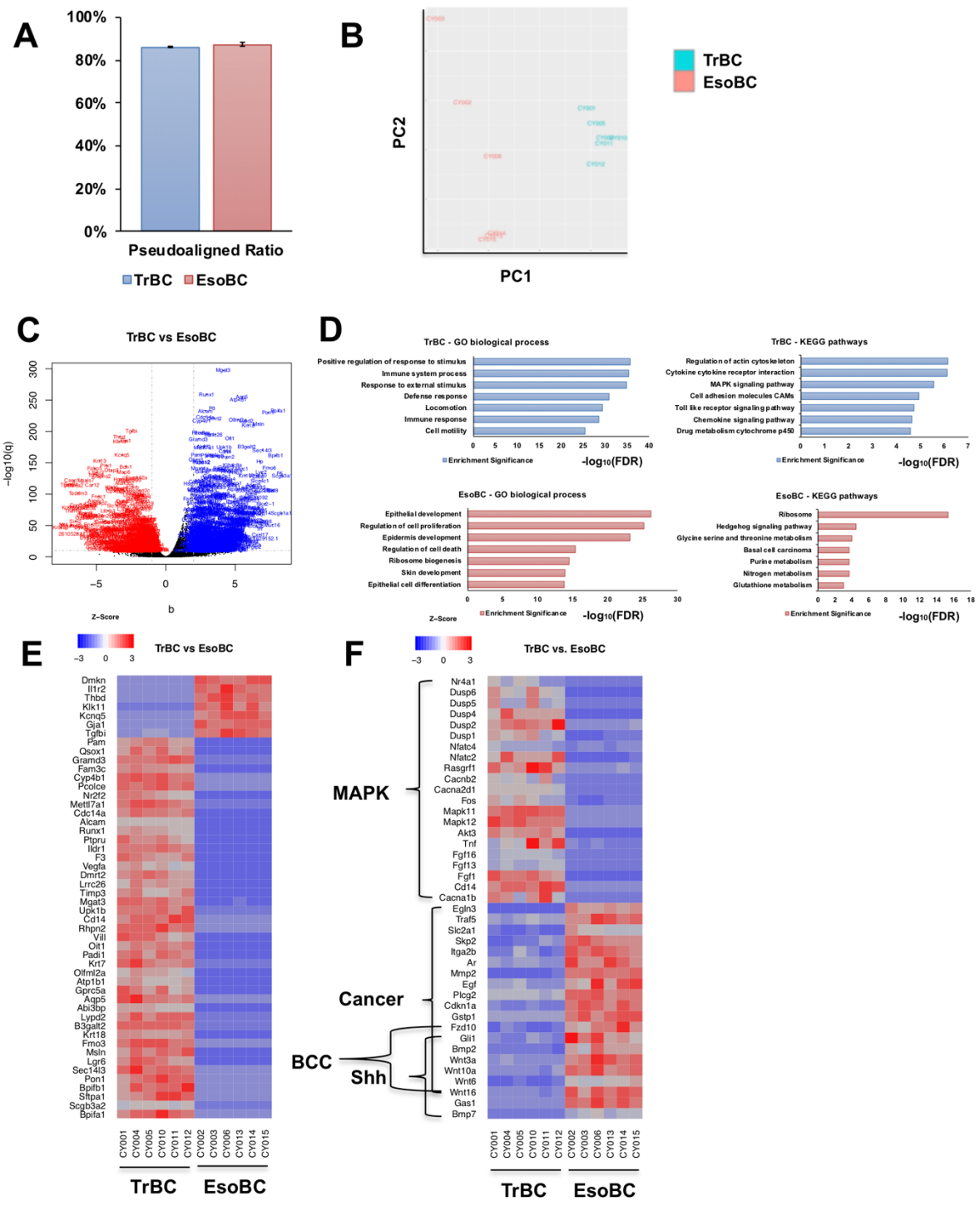


Figure 34. Transcriptome profiles of adult TrBCs and esoBCs.

(A) Bar graph showing the pseudoaligned ratios of adult TrBCs and esoBCs in an RNA-seq quantification program, Kallisto. **(B)** Principal component analysis (PCA) showing distinct difference between TrBC and esoBC samples. **(C**) Volcano plot showing differentially expressed genes in adult TrBCs and esoBCs. represented in black dots; genes significantly enriched in adult TrBCs were labeled with gene names in blue ($-\log_{10}(q)>10$; b>2); genes significantly enriched in adult esoBCs were labeled with gene names in red ($-\log_{10}(q)>10$; b<-1). **(D)** Enriched pathways in each population. Representative MSigDB gene sets (GO biological process & KEGG pathways) in adult TrBCs (blue) and adult esoBCs (red). **(E)** Top 50 genes ranked by significance showing dramatic difference in expression pattern in adult TrBCs and esoBCs. **(F)** Distinct expression pattern of genes in each representative KEGG pathways. **(E-F)** Relative expression levels of genes (row-wise Z-score of mean tpm). N=6 for each population.

Transcription factors significantly enriched in TrBCs included Nkx2-1, Hey1, Gata6, Epas1, Nr2f2. The over-representation of Foxj1, Foxi1, Ascl1 could be ascribed to early commitment of some TrBCs to differentiation programs towards multiciliated, ionocyte, and neuroendocrine lineages, respectively (Figure 35A). EsoBC-enriched transcription factors included Pitx1, Grhl3, Barx2, Gli1, Foxe1, Notch3, Ovol1, Zfp750, Notch1 (Figure 35B). The paired-like homeodomain transcription factor 1, Pitx1, was reported to play a pivotal role in promoting squamous carcinogenesis, cooperating with Sox2 and p63[133]. Foxe1, a downstream target of Shh pathway, was reported to be expressed in basal cell carcinoma[134]. Grhl3, Ovol1, Zfp750 were reported to be critical in epidermal development and squamous differentiation[135-137]. Notably, a number of activator protein-2 (AP2) transcription factors, TFAP2E, TFAP2C, TFAP2A, were specifically enriched in esoBCs but not in TrBCs. TFAP2C was suggested to function as an initiation factor, epigenetically inducing the surface ectoderm chromatin landscape in the early differentiation process of epidermis[79]. Furthermore, TFAP2C has been found to be overexpressed in squamous cell lung carcinoma, and to promote lung tumorigenesis and epithelial-mesenchymal transition[138]. Exploring the functional roles of these transcription factors may provide mechanistic insights into the regulatory networks governing tissue-specific BC differentiation potentials.

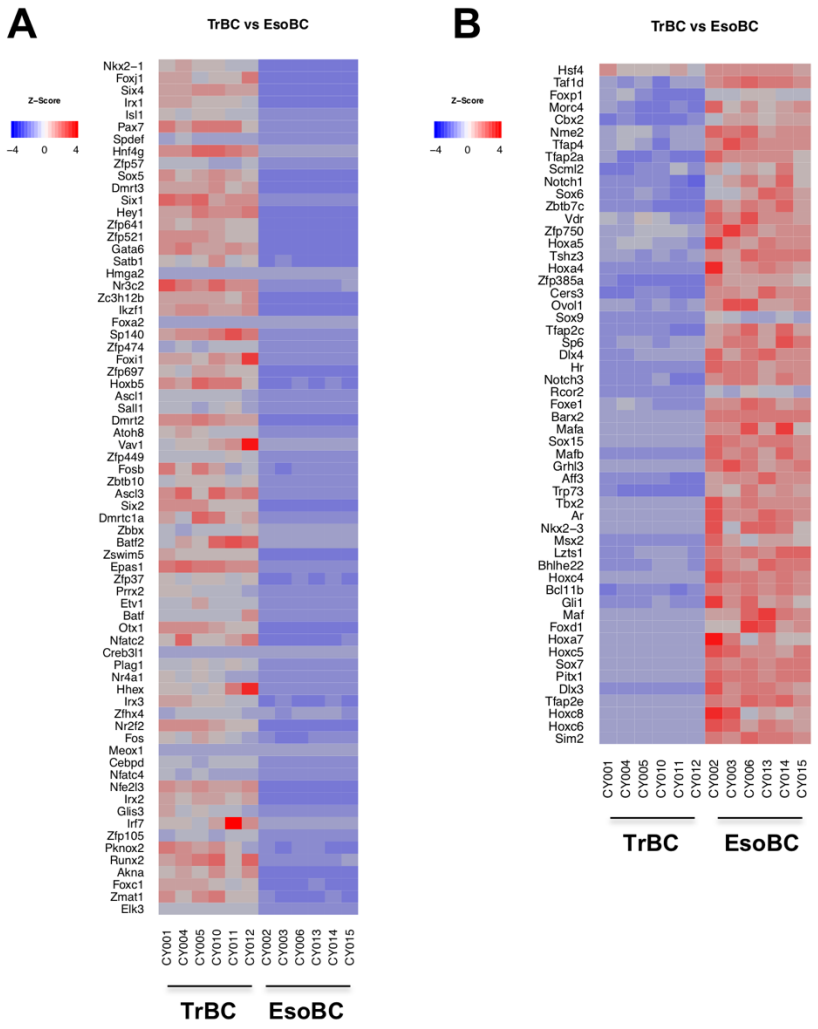


Figure 35. Distinct transcription factors enriched in adult TrBCs and esoBCs.

(A) Adult TrBC-specific transcription factors ($-\log_{10}(q)>10$; $b>2$). **(B)** Adult esoBC-specific transcription factors ($-\log_{10}(q)>10$; $b<-1$).

6-3. TrBC heterogeneity revealed by single cell RNA-seq analysis

Lineage tracing studies using *Krt5-CreERT2*[2] and *p63-CreERT2* support the idea that Krt5$^+$ p63$^+$ BCs are adult stem cells responsible for homeostatic maintenance and repair of the tracheal epithelium[139]. ~20% of TrBCs express Krt14, and previous studies have shown that this population was unipotent in homeostasis[6,140]. Additionally, unipotent BCs co-expressing low levels of c-myb or N2ICD were suggested to be primed for multiciliated or secretory lineage commitment, even before the acquisition of early luminal

progenitor marker Krt8[54]. A subsequent clonal lineage tracing study using *Krt5-CreERT2* and single cell qRT-PCR proposed that TrBCs include two subpopulations: basal stem cells (BSCs) expressing relatively higher levels of Dlk2, Dll1, Lmo1 and Snai2, and basal luminal precursors (BLPs) expressing relatively higher levels of Krt8 which labels cells that were still uncommitted but not stem-like as BSCs[4]. Interestingly, this study found Krt14 was equally detected in BSC and BLP populations, refuting the idea that Krt14$^+$ BCs was a unique subpopulation of BCs. However, single cell qRT-PCR could not provide a comprehensive gene expression analysis, and only 17 BCs were analyzed in this study.

To investigate the heterogeneity of TrBCs during homeostasis, 4,209 TrBCs from adult *p63-CreERT2; R26-tdTomato* mice (sorting strategy shown in Figure 31A) were isolated and profiled using 10X Genomics Chromium scRNA-seq. In total, 428 million reads were achieved (101,750 mean reads per cell), and the median value of the unique molecular identifier (UMI, each representing one transcript) number per cell was 10,802. 2,956 median genes were detected per cell (Figure 36A). Based on the whole transcriptomes, we identified 9 distinct clusters in these 4,209 cell profiles (Figure 36B). Markers such as EpCAM, Nkx2-1, Krt5 and Trp63 were widely expressed in all 9 clusters, confirming the purity of sorted TrBCs (Figure 36C & Figure 37). 243 cells expressing low levels of p63 and Krt5 and high levels of Scgb1a1 and Muc5b were annotated as Secretory cluster. 139 cells were considered as Squamous cluster as they expressed Krt14 and Krt13. 96 cells co-expressing EpCAM, Nkx2-1, Krt5, Trp63 and Vim (Vimentin) were annotated as EMT cluster (epithelial-mesenchymal transition). 90 cells partially overlapped with the Krt14$^+$ population expressing cell cycle markers, Mki67, Nusap1, Aurkb, were identified as Proliferating cluster (2.14% of total TrBC population). 28 Foxj1$^+$ multiciliated cells and one Ascl1$^+$ neuroendocrine cell were also identified. 4 cells were annotated as Immune-related cluster as they co-expressed EpCAM, Nkx2-1, Krt5, and some immune cell markers CD244 (Natural Killer Cell receptor 2B4), Irf8 (Interferon Regulatory Factor 8), CD74 (HLA class II histocompatibility antigen gamma chain) and others. The majority of the cells were clustered into two BC subpopulations, Basal-1 (1,939 cells) and Basal-2 (1,669 cells) (Figures 36B-C & Figure 37).

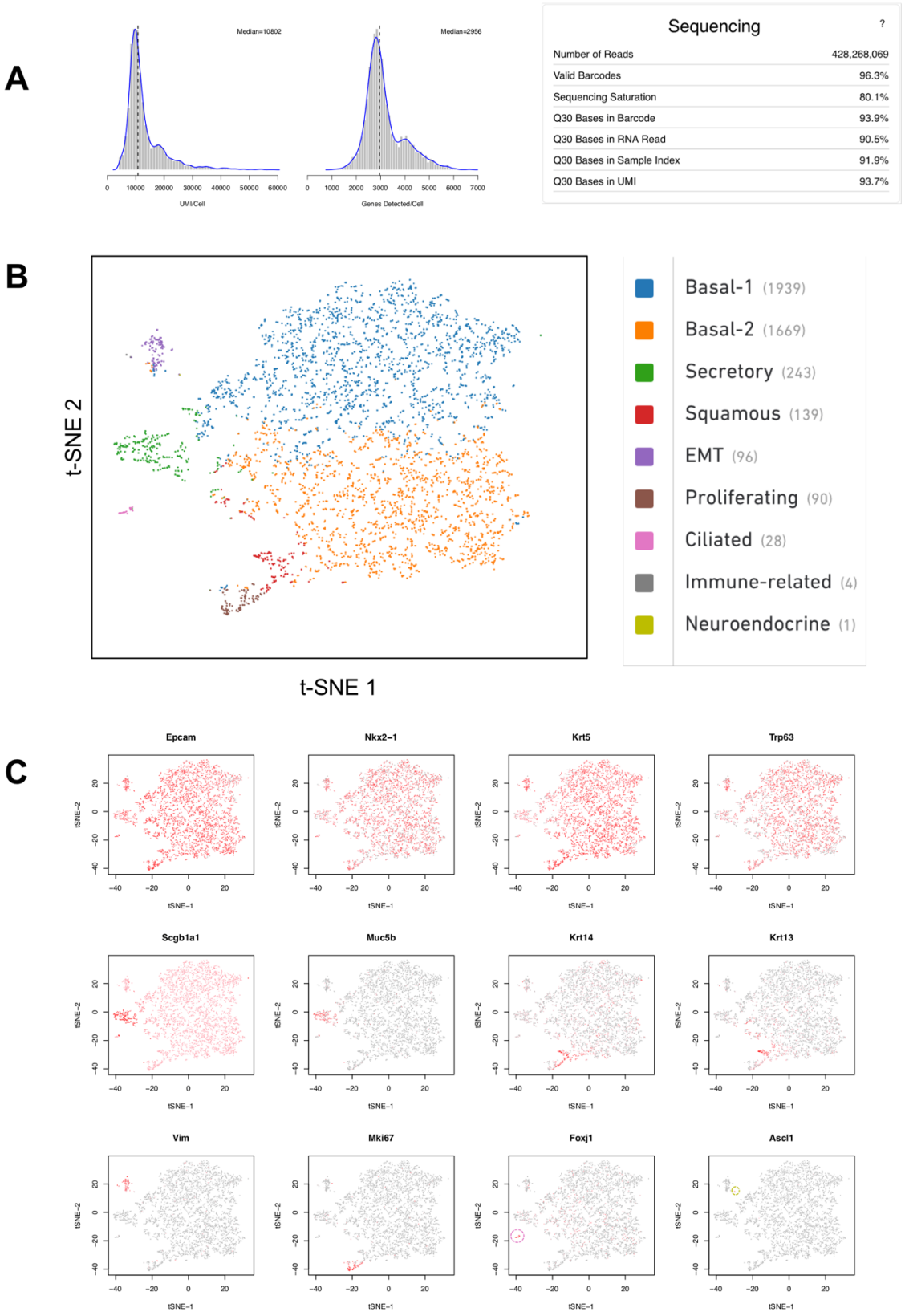


Figure 36. Clustering of tracheal BCs by single-cell RNA sequencing.

(A) Quality control for the 10X Genomics scRNA-seq data. Distributions of the number of UMI per cells (left), the number of genes (middle), and the sequencing information (right). **(B)** t-distributed stochastic neighbor embedding (t-SNE) of 4,209 scRNA-seq profiles, colored by cluster assignment and annotated using known lineage-specific markers. **(C)** *Post hoc* cluster interpretation based on expression of known lineage-specific markers. t-SNE presentation of 4,209 cells, colored by expression (log_2(TPM+1)) of each marker gene.

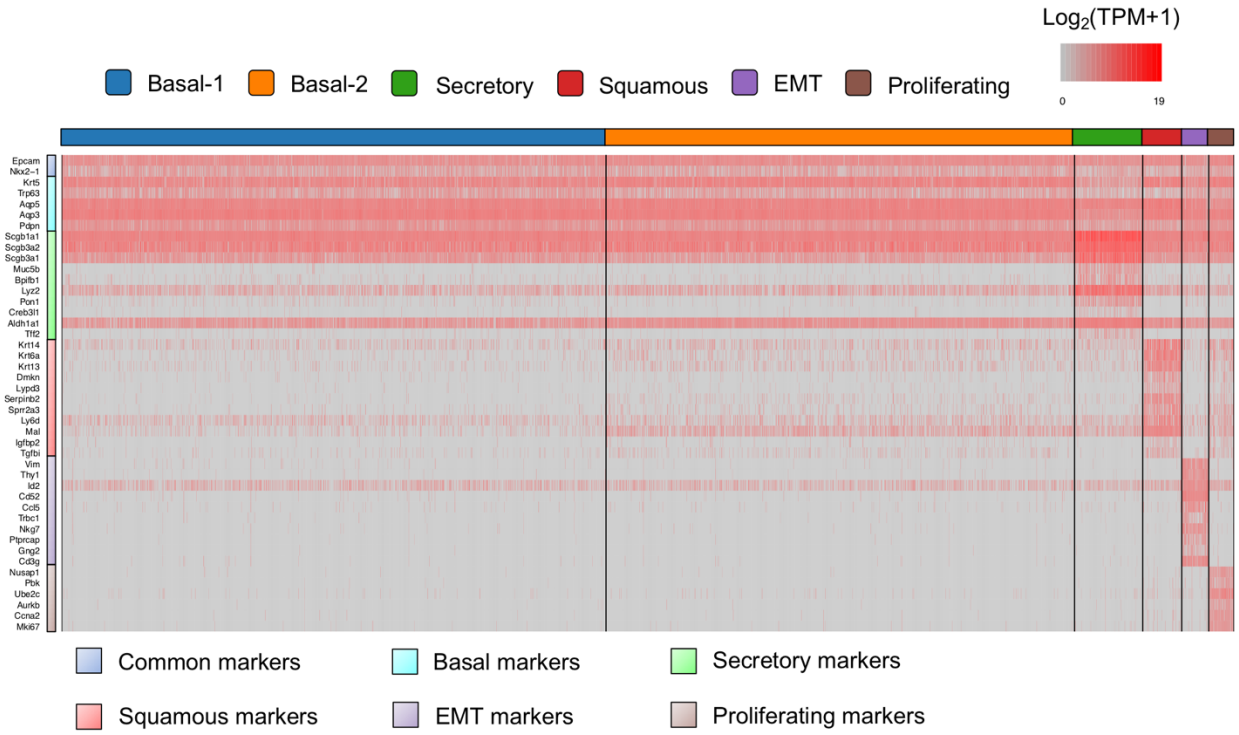


Figure 37. Expression levels (log_2(TPM+1)) of cluster-distinct genes (rows) in each epithelial cells (columns).

By analyzing the expression pattern of cluster-specific genes, we found that many secretory genes, including Scgb1a1, Scgb3a2, Scgb3a1 Lyz2, Aldh1a1 were widely expressed in all clusters with slightly higher expression levels in the Secretory cluster. In contrast, Muc5b, Bpifb1, Pon1, Creb3l1 and Tff2 were considered as more specific secretory lineage markers or representing a more mature status for secretory lineage (Figure 37).

The scRNA-seq profiles suggested two major clusters Basal-1 and Basal-2, with almost 1:1 ratio in the TrBCs. To further understand about the nature of these two TrBC subpopulations, their transcriptomes were compared (Figure 38A). Gene set overlapping analysis revealed that Basal-1 cluster cells expressed genes associated with drug metabolism, glutathione metabolism and retinol metabolism, while Basal-2 cluster cells expressed genes associated with WNT, TGFB and Hedgehog signaling pathways, non-small cell lung cancer (NSCLC), basal cell carcinoma (BCC) (Figure 38B). Interestingly, Hedgehog pathway, NSCLC, BCC had been found associated with genes specifically expressed in esoBCs compared to TrBCs (Figures 34D & F). Cited2 (Cbp/p300 interacting transactivator with Glu/Asp-rich carboxy-terminal domain 2), a transcriptional modulator, was more enriched in Basal-1 (Figures 38C&D). Cyp2b10 was more enriched in Basal-2, suggesting potentially a more differentiated subpopulation comparing to Basal-1. Tppp3 (tubulin polymerization-promoting protein family member 3), the most enriched gene in Basal-2, was reported to act as an oncogene in NSCLC[141]. Another Basal-2 enriched gene, Cav1, was also reported to be required for tumor cell survival and growth in NSCLC[142] (Figure 38C). EsoBC-specific transcription factors, Barx2, Klf4, were preferentially expressed in Basal-2, suggesting that Basal-2 might represent a cellular status more similar to BCs in stratified squamous epithelium (Figures 38C &D). Intriguingly, Isl1 is expressed in the ventral mesenchymal and epithelial cells of anterior foregut during development[143], while Srf is a key smooth muscle differentiation gene expressed in dorsal mesenchymal compartment of trachea during development[144]. The enrichment of Isl1 in Basal-1 and Srf in Basal-2 may give some hints on the potential distinct spatial locations of these two subpopulations in the dorsal-ventral axis (Figure 38D).

The 1:1 ratio of Basal-1 and Basal-2 subpopulations identified in the scRNA-seq analysis was reminiscent of the BSCs and BLPs proposed in Watson et al., 2015[125]. To examine the correlation of Basal-1/Basal-2 and BSC/BLP, the distinct markers identified in the previous study which could distinguish BSCs and BLPs were analyzed again here. Krt8 was shown to be more highly expressed in BLPs, while Dlk2, Dll1, Lmo1, Snai2 were more enriched in BSCs[4]. In the differential gene expression analysis between Basal-1 and Basal-2, Krt8 was more highly expressed in Basal-2 (p=5.18E-139; d=-1.73); Dlk2 was more enrich in Basal-1 (p=1.38E-13; d=0.92); Dll1 and Lmo1 was not differentially expressed (Dll1: p=0.78; d=-0.03; Lmo1: p=0.25; d=0.13); Snai2 was slightly enriched in Basal-2 (p=0.001; d=-0.34). Due to limited markers presented in the previous paper, the similarity between Basal-1/Basal-2 and BSC/BLP could not

be confidently assessed. Nevertheless, as discussed above and following, the Basal-2 subpopulation might represent a more luminal committed BC cluster with higher potential for squamous differentiation, and may assume the dorsal location adjacent to esophagus.

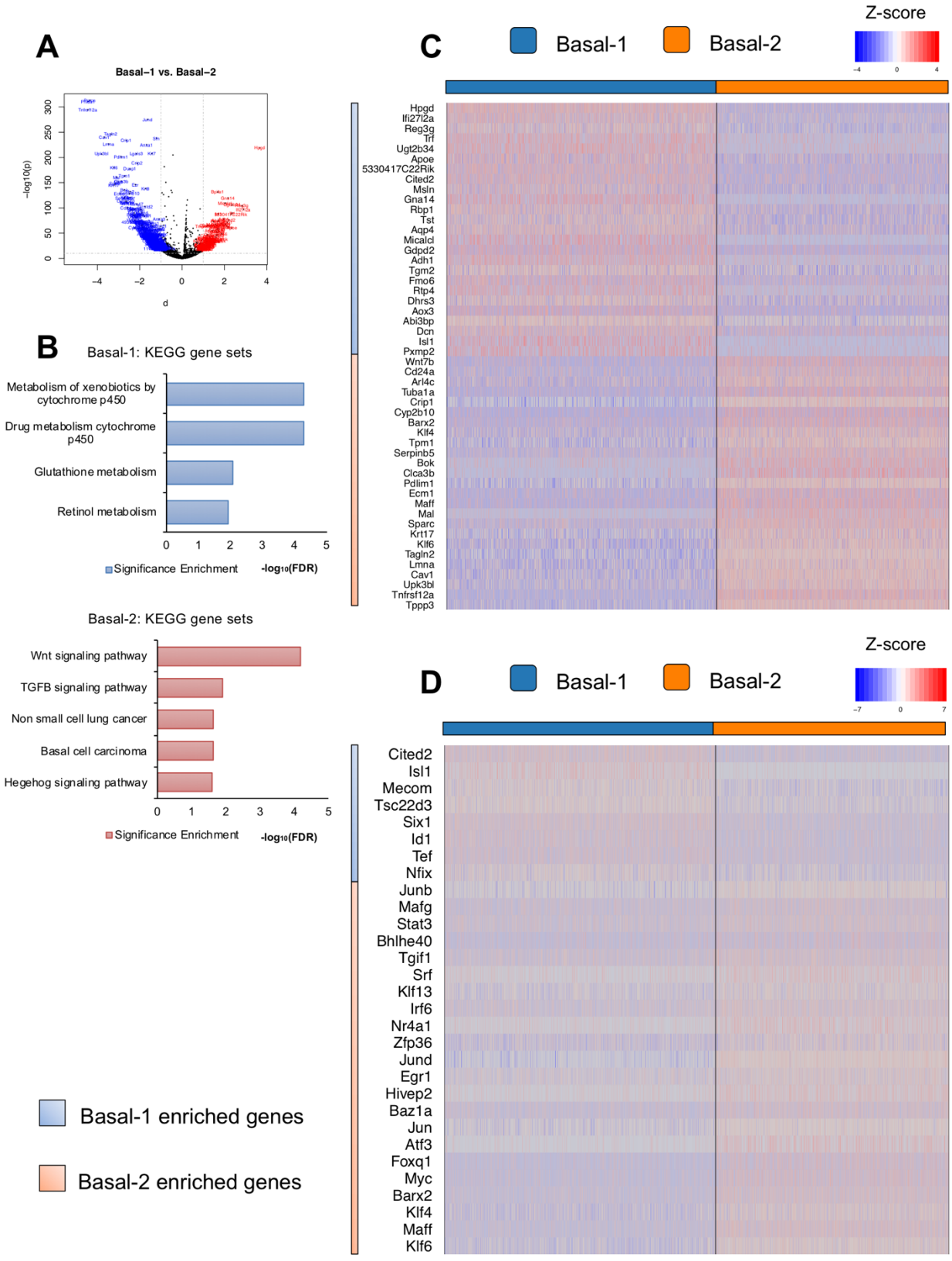

A
Basal−1 vs. Basal−2
−log10(p)
d
B
Basal-1: KEGG gene sets
Metabolism of xenobiotics by cytochrome p450
Drug metabolism cytochrome p450
Glutathione metabolism
Retinol metabolism
Significance Enrichment
-log10(FDR)
Basal-2: KEGG gene sets
Wnt signaling pathway
TGFB signaling pathway
Non small cell lung cancer
Basal cell carcinoma
Hegehog signaling pathway
Significance Enrichment
-log10(FDR)
Basal-1 enriched genes
Basal-2 enriched genes
C
Basal-1
Basal-2
Z-score
Hpgd
Ifi27l2a
Reg3g
Trf
Ugt2b34
Apoe
5330417C22Rik
Cited2
Msln
Gna14
Rbp1
Tst
Aqp4
Micalcl
Gdpd2
Adh1
Tgm2
Fmo6
Rtp4
Dhrs3
Aox3
Abi3bp
Dcn
Isl1
Pxmp2
Wnt7b
Cd24a
Arl4c
Tuba1a
Crip1
Cyp2b10
Barx2
Klf4
Tpm1
Serpinb5
Bok
Clca3b
Pdlim1
Ecm1
Maff
Mal
Sparc
Krt17
Klf6
Tagln2
Lmna
Cav1
Upk3bl
Tnfrsf12a
Tppp3
D
Basal-1
Basal-2
Z-score
Cited2
Isl1
Mecom
Tsc22d3
Six1
Id1
Tef
Nfix
Junb
Mafg
Stat3
Bhlhe40
Tgif1
Srf
Klf13
Irf6
Nr4a1
Zfp36
Jund
Egr1
Hivep2
Baz1a
Jun
Atf3
Foxq1
Myc
Barx2
Klf4
Maff
Klf6

Figure 38. BC subtypes showing unique gene expression patterns.

(A) Volcano plot showing differentially expressed genes in Basal-1 and Basal-2 clusters represented in black dots; genes significantly enriched in Basal-1 were labeled with gene names in red ($-\log_{10}(q)>10$; $d>1$); genes significantly enriched in adult Basal-2 were labeled with gene names in blue ($-\log_{10}(q)>10$; $d<-1$). **(B)** Enriched KEGG gene sets in each population in Basal-1 (blue) and Basal-2 (red). **(C)** Top 25 genes significantly enriched in Basal-1 and Basal-2 showing markedly difference in expression pattern in these two clusters. **(D)** Basal-1 and Basal-2 specific transcription factors. **(C-D)** Relative expression levels of genes (row-wise Z-score of mean $\log_2(TPM+1)$).

The Squamous cluster was intriguing, as normal pseudostratified tracheal epithelium could potentially undergo metaplastic transformation into squamous epithelium in injury challenge or pathological conditions. To characterize this subpopulation, differential gene expression profiles were generated by comparing the Squamous cluster to all the other cells (Figure 39A). Genes significantly increased in the Squamous cluster were associated with epidermal development, keratinocyte differentiation, keratinization, while genes significantly decreased in this cluster was associated with different metabolic processes and cellular homeostasis (Figure 39B). Gene signature for this cluster was generated, shown by top 50 genes with most differentially expressed pattern (Figure 39C). A recent scRNA-seq study revealed a cell population with high Krt13/Krt4 expression and named as hillock cells. These cells were reported to form stratified squamous structure in contiguous groups[13]. By crossing the gene signature of the Squamous cluster in this analysis with the hillock cell signature, 15 out of 50 top differentially expressed genes in the hillock cells were found overlapping with the top 50 signature genes in our Squamous cluster. The possibility could not be ruled out that the basal subpopulation in the hillock cells might be the same as this Squamous cluster in the present study. Many secretory cell-related genes such as Lyz2, Cyp2a5, Aldh1a1, Scgb1a1 were overrepresented in the hillock cell-associated genes, which was the weakness of Montoro et al., 2018[13] ("hillock" cells were defined as $Krt13^+$ $Krt4^+$ cells, containing both the minor subpopulation of $Krt5^+$ BCs and the major subpopulation of $Krt5^-$ LCs). The transcription factors enriched in the Squamous cluster included Barx2, Klf4, and Pax9, all associated with esoBC features or squamous differentiation. From the heatmap, the expression levels of these three key genes were found to increase from Basal-1 to Basal-2 to Squamous

cluster. Trp63 was significantly decreased in the Squamous cluster, indicating a relatively more differentiated status towards luminal lineages. Tsc22d3 was the most significantly decreased transcription factor in the Squamous cluster (Figure 39D). Interestingly, this gene was more highly expressed in Basal-1 compared to Basal-2 (Figure 38D), further suggesting a trend for the squamous differentiation potentials increasing from Basal-1 to Basal-2 to Squamous cluster.

To our knowledge, the transcriptome analysis here provided the first comprehensive characterization of TrBC heterogeneity in homeostasis with high resolution at gene expression level. In the recent two scRNA-seq studies, several BC markers were identified[13,145]. By assessing the expression pattern of these BC markers, majority of them were found failed to distinguish the TrBC clusters in the present dataset: Hlf, Icam1, Notch1, Ngfr, Krt15, Krt5, F3, Aqp3 were widely distributed in all clusters; Epas1 was more enriched in Secretory cluster; Procr, Sfn were more highly expressed in Squamous cluster. The signatures for each TrBC clusters identified in the present study generated novel markers for future *in situ* validation. Furthermore, the gene set overlapping analyses of distinct clusters revealed a trend towards squamous differentiation potential in different TrBC clusters. Further functional characterization of these genes could provide novel mechanistic insights into the pathogenesis of squamous metaplasia and squamous carcinoma.

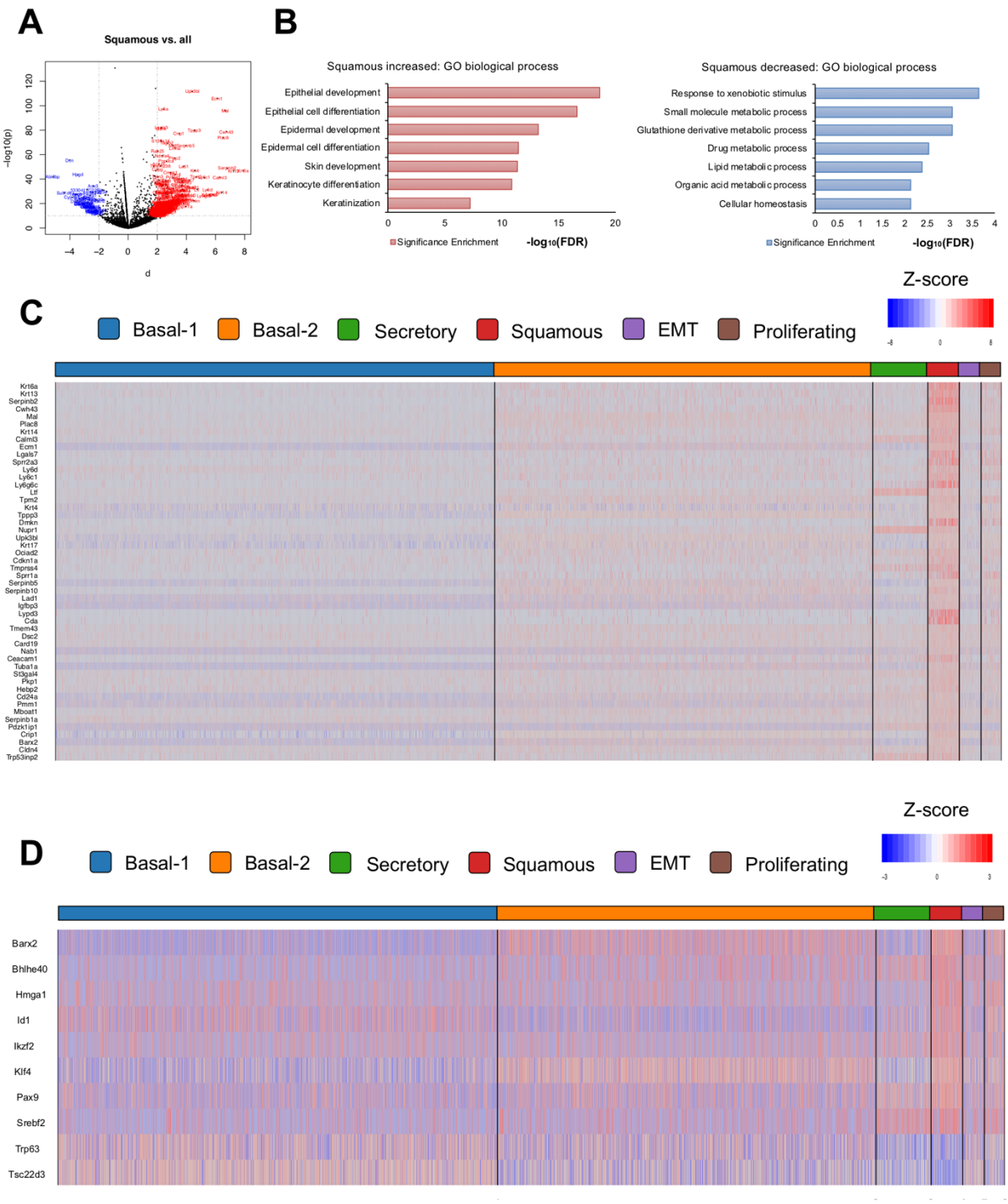


Figure 39. Squamous cluster showing unique gene expression patterns comparing to the rest cells.

(A) Volcano plot showing differentially expressed genes in Squamous cluster and the rest cells represented in black dots; genes significantly enriched in Squamous cluster were labeled with gene names in red ($-\log_{10}(q)>10$; $b>2$); genes significantly decreased in Squamous cluster were labeled with gene names in blue ($-\log_{10}(q)>10$; $b<-2$). **(B)** Enriched KEGG gene sets in genes significantly increased in the Squamous cluster (red) and significantly decreased in the Squamous cluster (blue). **(C)** Squamous cluster gene signature. Top 50 genes ranked by fold change significantly enriched in Squamous cluster comparing to the rest cells. **(D)** Squamous cluster specific transcription factors. **(C-D)** Relative expression levels of genes (row-wise Z-score of mean $\log_2(TPM+1)$).

6-4. EsoBC heterogeneity revealed by single cell RNA-seq analysis

The mouse esophagus is lined by a stratified squamous epithelium, comprised of $EpCAM^+$ $Sox2^+$ $p63^+$ BCs, $EpCAM^-$ $Sox2^-$ suprabasal cells and a keratinized superficial layer. The esophageal epithelium is surrounded by a relatively homogenous stromal and muscle layer, leading to the hypothesis that esoBCs are homogenous and equipotential, similar to the interfollicular basal cells in epidermis[146]. Two subsequent studies suggested that esoBCs are heterogeneous. One study using *in vitro* organoid culture and FACS analysis proposed that in the esoBC population, Itga6/$Itgb4^{high}$ $CD73^+$ cells represent a less differentiated non-quiescent stem cell subpopulation, while the majority serve as transit-amplifying cells. Itga6/$Itgb4^{high}$ $CD73^-$ cells are considered as early transit-amplifying cells, and Itga6/$Itgb4^{low}$ cells are late transit-amplifying cells[147]. The second study, using lineage tracing, showed that $Krt15^+$ progenitors are the long-lived stem cell population responsible for homeostasis and injury repair[148]. However, none of these studies addressed the heterogeneity from an unbiased way using comprehensive gene expression analysis.

To investigate esoBC heterogeneity, 10X Genomics Chromium scRNA-seq was performed in 4,383 sorted esoBCs from adult mice under homeostatic conditions (sorting strategy in Figure 33A), generating 343 million reads in total, 17,815 median UMI per cell, and 3,504 median genes per cell (Figure 40A). Based on the whole transcriptomes, 4 distinct clusters were identified for these 4,383 profiles (Figure 40B). 12 cells were considered as contaminated Mesenchyme, as they lacked EpCAM, Sox2, Krt5, Trp63 but expressed Vim. All the rest expressed EpCAM, Sox2, Krt5 and p63, confirming the purity of sorted esoBCs. 582 cells were annotated as Proliferating cluster as they expressed Mki67, Nusap1, Aurkb and other cycling

markers (Figures 40B-C, Figure 41). This Proliferating cluster constituted 13.2% of total esoBC population. 3,082 cells were identified as Basal-1 cluster while the other 707 cells were annotated as Basal-2, to distinguish their relative differentiation status (Figure 40B-C). Tgfbi and Dlk2 were biased enriched in Basal-1 cluster, while genes associated with more squamous differentiation status, Krt13, Grhl1, Krt6a, and others. were biased enriched in Basal-2 cluster (Figure 40C & Figure 41). Interestingly, Dlk2 was found to be more enriched in Basal-1 cluster in TrBC dataset. This led to a hypothesis that Dlk2 marked a relatively less differentiated, more stem-like BC subpopulation in both trachea and esophagus. Dlk2 functions as an inhibitory non-canonical ligand for Notch1 receptor. Notch1 was expressed in both TrBCs and esoBCs. This suggested that the Notch activity might be repressed autonomously in the TrBC Basal-1 cluster and the esoBC Basal-1 clusters through the expression of Dlk2, consistent with the observation that pharmacological Notch inhibition led to expansion of TrBCs[3,122] and Notch1 activation was required for the differentiation of esoBCs towards squamous lineage[149].

Previous studies suggested that proliferative activity was equally possessed by different esoBC subpopulations, thus leading to the theory of equipotent population-level stem cell activity[147]. The scRNA-seq data presented here revealed that at least from the gene expression level, the Proliferating cluster was remarkably distinct from other clusters (Figure 41).

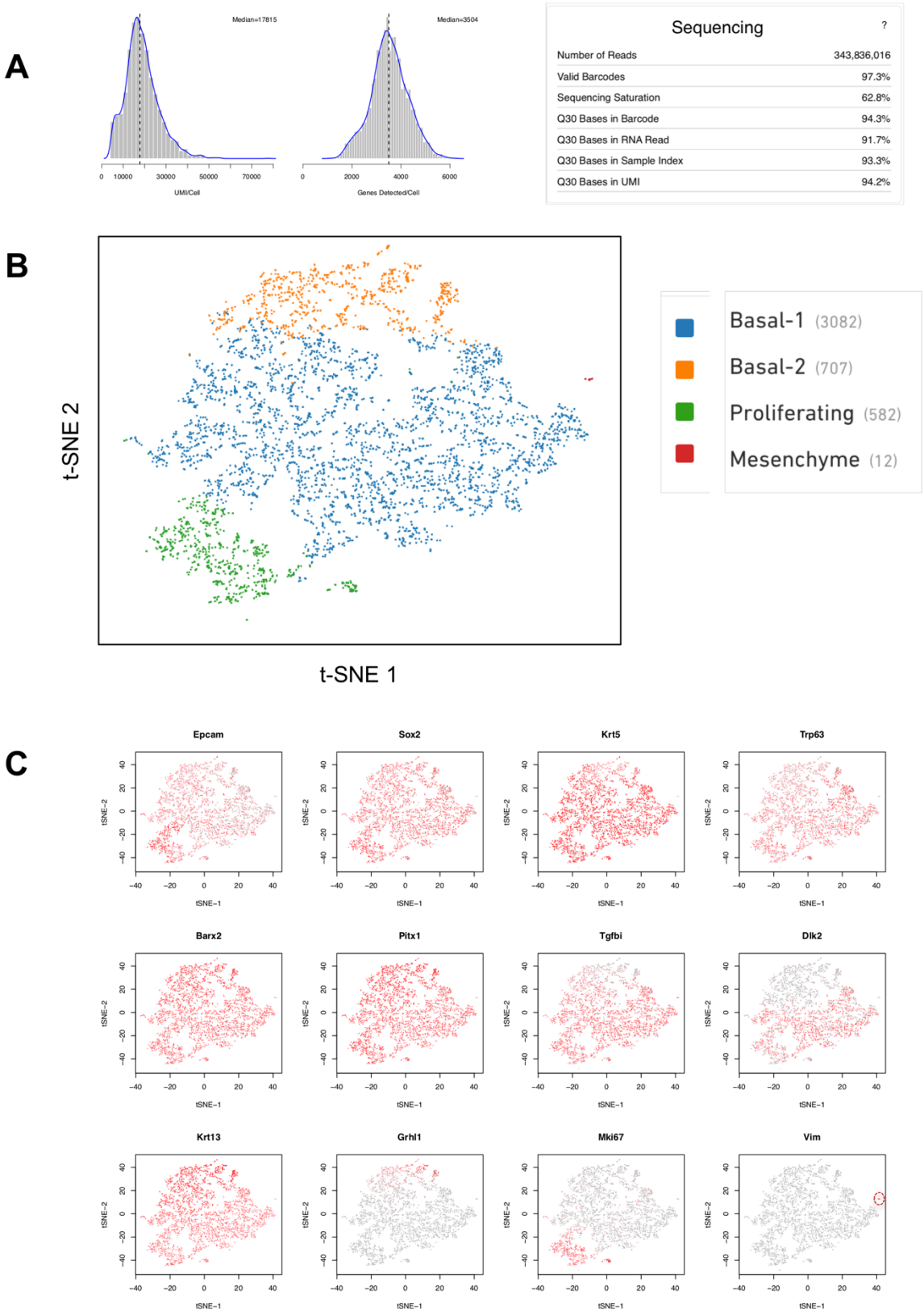


Figure 40. Clustering of esophageal BCs by single-cell RNA sequencing.

(A) Quality control for the 10X Genomics scRNA-seq data. Distributions of the number of UMI per cells (left), the number of genes (middle), and the sequencing information (right). **(B)** t-distributed stochastic neighbor embedding (t-SNE) of 4,383 scRNA-seq profiles, colored by cluster assignment and annotated using known lineage-specific markers. **(C)** *Post hoc* cluster interpretation based on expression of known lineage-specific markers. t-SNE presentation of 4,383 cells, colored by expression (log_2(TPM+1)) of each marker gene.

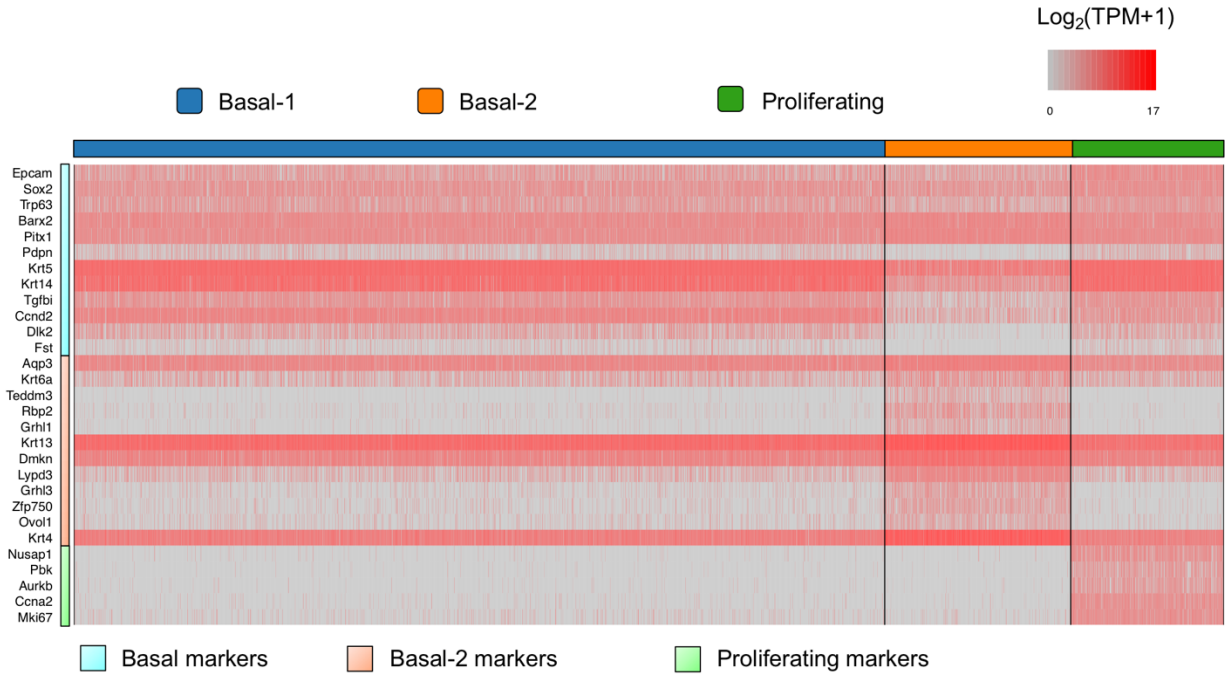


Figure 41. Expression levels (log_2(TPM+1)) of cluster-distinct genes (rows) in each cells (columns).

To assess the differences between Basal-1 and Basal-2 clusters, differential gene expression analysis was performed for these two clusters (Figure 42A). Consistent with the expectation that Basal-1 cluster represented a less differentiated cell status and Basal-2 cluster represented a more luminal committed differentiation status, the gene ontogeny analysis revealed that Basal-1 cluster-specific genes were more associated with hemidesmosome assembly, organ morphogenesis, regulation of cell proliferation, response to growth factor, while the Basal-2 cluster-specific genes were more overlapped with biological processes including lipid metabolic process, epidermis development, keratinocyte differentiation (Figure 42B). In

contrast to DeWard et al., 2014[147], no biased expression pattern for CD73 was observed. Nevertheless, Krt15 was indeed more enriched in the Basal-1 cluster (p=2.15E-132; d=2.78), consistent with the previous report that Krt15^{+} subpopulation behaved more as progenitor-like cells[148]. Interestingly, based on the current analysis, Krt14 (p=7.66E-205; d=4.77) compared to Krt15 was more reliable to distinguish Basal-1 versus Basal-2 clusters. The top gene most enriched in Basal-1 cluster, Wfdc2 (WAP four-disulfide core domain 2), was highly expressed in all 9 clusters from TrBC dataset, and had been associated with immune defense response in respiratory system[150]. Overexpression of Wfdc2 in endometrial cancer cell lines led to enhanced cell proliferation and colony formation[151]. The functional significance of this highly biased expression of Wfdc2 in esoBC Basal-1 cluster required further investigation.

The specific enrichment of Wnt10a (p=3.12E-204; d=4.72), Igfbp7 (p=2.18E-170; d=4.65), Tgfbi (p=3.16E-143; d=3.92) in the Basal-1 cluster underpinned the importance of WNT, IGF, TGFB signaling pathways in maintenance of the esoBC progenitor potential, awaiting functional validation in the future (Figure 42C). Krtdap was the top gene significantly enriched in the Basal-2 cluster (Figure 42C). Consistently, Krtdap was found to mark the early differentiating keratinocyte in mouse oral epithelium[152]. Grhl1, Grhl3, Notch3, Klf4 were significantly enriched in Basal-2 cluster, representing a more squamous differentiated activity (Figure 42D).

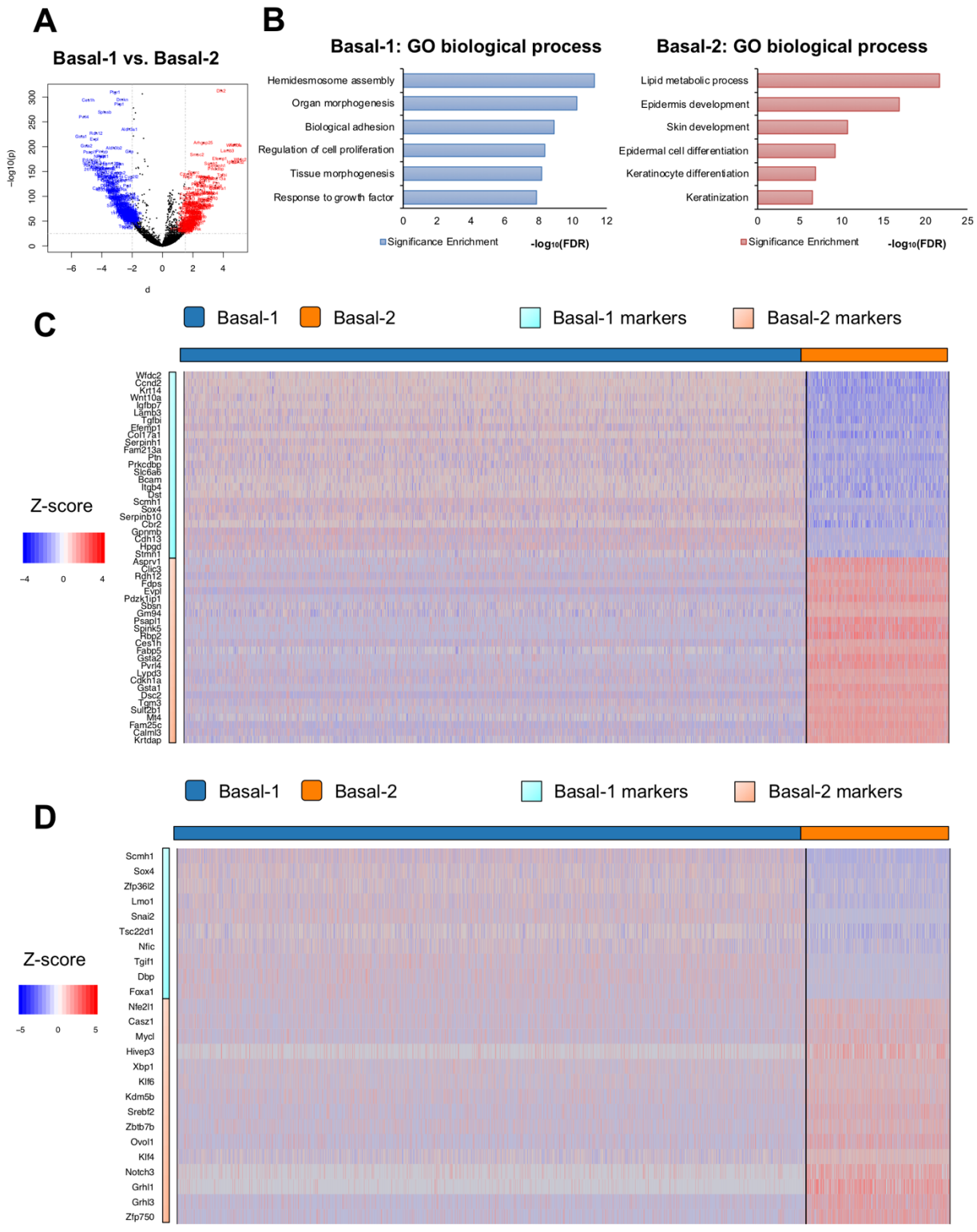


Figure 42. BC subtypes showing unique gene expression patterns.

(A) Volcano plot showing differentially expressed genes in Basal-1 and Basal-2 clusters represented in black dots; genes significantly enriched in Basal-1 cluster were labeled with gene names in red ($-\log_{10}(q)>25$; $d>1.5$); genes significantly enriched in Basal-2 cluster were labeled with gene names in blue ($-\log_{10}(q)>25$; $d<-2$). **(B)** Enriched GO biological process in Basal-1 cluster (blue) and Basal-2 cluster (red). **(C)** Top 25 genes significantly enriched in Basal-1 and Basal-2 clusters showing dramatic difference in expression pattern in these two clusters. **(D)** Basal-1 cluster and Basal-2 cluster specific transcription factors. **(C-D)** Relative expression levels of genes (row-wise Z-score of mean $\log_2(TPM+1)$).

To further investigate the three distinct esoBC clusters, the esoBC scRNA-seq data was interrogated again using "virtual inference of protein activity by enriched regulon analysis" (VIPER), a robust and unbiased systems biology algorithm to assess protein activity from gene expression data[113]. Tissue-specific interactomes were built through the algorithm for the reconstruction of accurate cellular networks (ARACNe[153]) using the 4,383 esoBC scRNA-seq profiles. In general, ARACNe first identifies statistically significant regulations through mutual information (MI), and then eliminates indirect regulations through data processing inequality (DPI)[112,153]. ARACNe networks are then used in the VIPER algorithm as the basis for protein activity assessment. Conventional gene expression profile analysis uses the expression levels of each single gene. However, scRNA-sequencing technology suffers from inevitable technical noises, mainly by low sequencing depth and large gene dropout effects. More adversely, genes, especially transcription factors, which are expressed at relatively low levels, have a lower chance to be detected in general. In contrast, ARACNe-VIPER algorithms assess the protein activity of upstream regulators using the gene expression pattern of a battery of downstream targets, thus more resistant to the technical noises caused by gene dropout effects and more suitable to interrogate scRNA-seq datasets. Here, protein activities for each single cell were scored through VIPER, and the 4,383 esoBCs were presented in multidimensional scaling (MDS) plot based on the protein activities of VIPER inference. Consistent with the gene expression analysis, three discrete clusters could be identified here (Figure 43A).

IterClust, an unsupervised statistic framework for iterative clustering analysis[115], was then used to elucidate the heterogeneity of esoBCs in a hierarchical way. Round 1 separated C1-1 and C1-2 (Figure 43B left panel), and Round 2 further separated the C1-1 in Round1 into 3 new subclusters (C2-1, C2-2, C2-

3) and the C1-2 in Round 1 into 2 new subclusters (C2-4 and C2-5) (Figure 43B right panel). Based on *post hoc* annotation, C2-4 was found to represent proliferating cells, not only showing exclusive expression of multiple cycling markers such as Ccna2, Hmgb2, Cenpf, Cenpk, Rbl1, but exhibiting remarkably high protein activity scores (Figure 43C top panel and data not shown). The C2-5 represented bona fide basal progenitor cells. Even though the expression levels of p63, Sox4, Sox15, Pax9, Foxa1 did not significantly differ among these clusters, the protein activity was much higher in C2-5 and was negatively presented in C1-1, which further demonstrated that the protein activity analysis outcompeted gene expression analysis in this case. Another progenitor marker, Snai2, showed both enriched expression and higher protein activity in C2-5 (Figure 43C middle panel and data not shown). The C1-1 was considered as more luminal differentiated cellular status, as Notch3, Zfp750, Grhl1, Grhl3, Ovol1 were significantly more expressed in C1-1 and showed higher protein activities in C1-1 (Figure 43C bottom panel and data not shown). Another differentiation gene, Pycard, showed comparable expression levels in all clusters, but strong positive protein activity in C1-1 and negative protein activity in C2-4 and C2-5 (data not shown).

This iterClust analysis suggested that the most pronounced differences in esoBC population was between the less differentiated cells (C2-5 and C2-4) and the more differentiated cells (C1-1). The difference between the progenitor cluster C2-5 (similar to the Basal-1 cluster in the expression analysis above) and the proliferating cluster C2-4 (similar to the Proliferating cluster in the expression analysis above) was subtler than the difference in lineage progression. Again, the esoBCs were found to consist of three major subpopulations, one in cycling status, one with more progenitor features, and one more committed towards luminal differentiation. The additional segregation of more differentiated C1-1 cells into 3 subclusters in Round 2 (C2-1, C2-2, C2-3) might suggest three distinct lineage-primed differentiation programs in these cells, reminiscent of the lineage segregation theory in tracheal basal compartment proposed in Pardo-Saganta et al., 2015[54]. This however needs further characterization and validation.

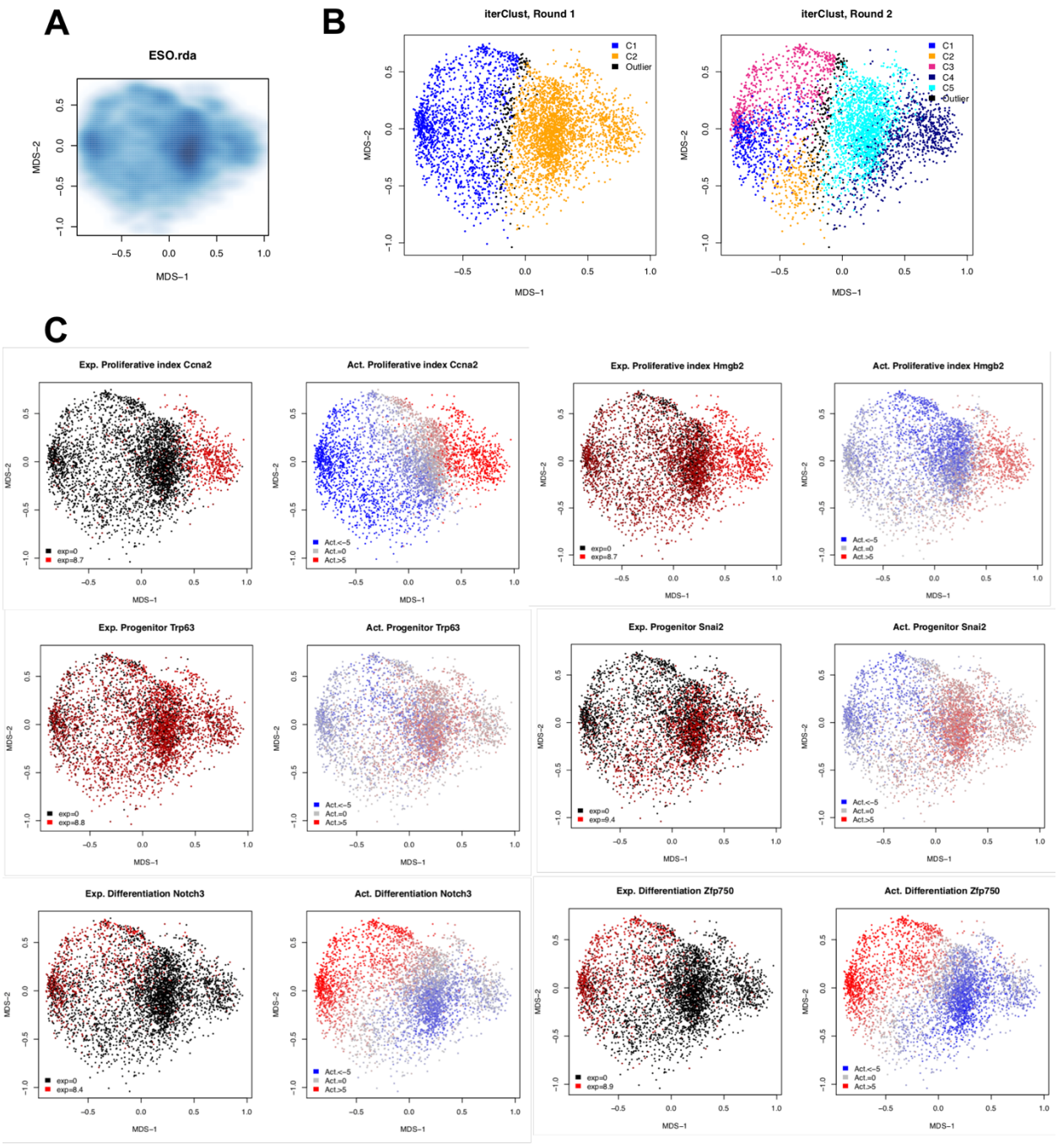


Figure 43. Three subpopulations of esoBCs revealed based on protein activity profiles

(A) Multidimensional scaling (MDS) visualization of 4,383 esoBCs based on protein activity analysis. **(B)** Iterative clustering analysis (iterClust) revealing three major clusters in esoBCs (C2-5, C2-4, C2-1+C2-2+C2-3 from Round 2). **(C)** *Post hoc* cluster annotation by the expression and activity of known markers. MDS presentation of 4,383 scRNA-seq profiles (each dot), colored by the expression level (log_2(TPM+1))

of each marker gene in the expression plots, or the protein activity level (inferred through the ARACNe-VIPER pipeline) of each marker protein in the activity plots.

6-5. A subpopulation of TrBCs exhibiting esoBC-like features

Trachea and esophagus share same developmental origin from the anterior foregut endoderm. The dorsal-ventral patterning specifies ventral tracheal as $Sox2^+$ $Nkx2.1^+$ and dorsal esophagus as $Sox2^+$ $Nkx2.1^-$. BCs from trachea and esophagus share common BC markers such as Trp63, Krt5, and they both serve as adult stem cells for homeostatic maintenance and injury repair. However, these two BC populations have intrinsic differences in self-renewing activity and differentiation programs: TrBCs are relatively quiescent, with differentiation potentials towards luminal secretory (Scgb3a2, Scgb1a1) and multiciliated (Foxj1) lineages to maintain the pseudostratified epithelial architecture. In contrast, esoBCs are highly active in proliferation to fulfill the high turn-over rate of esophageal epithelium, and generate $Krt4^+$ $Krt13^+$ luminal descendants to maintain the stratified squamous epithelial architecture.

In pathological conditions such as COPD (chronic obstructive pulmonary diseases), TrBCs may adopt an esophageal like differentiation program and generate squamous epithelium. The squamous metaplasia is reversible. Intriguingly, this pathological conversion can be recapitulated in air-liquid interface cultures. Under normal culture conditions, TrBCs generate pseudostratified epithelium with luminal secretory and multiciliated descendants, while esoBCs generate squamous stratified epithelium with $Krt4^+$ $Krt13^+$ suprabasal cells. However, upon withdrawal of retinoic acid, TrBCs can generate esophageal-like squamous stratified epithelium *in vitro*.

The scRNA-seq analyses of TrBCs revealed that some subpopulations of TrBCs have features associated with squamous differentiation (Basal-2 & Squamous clusters in TrBC dataset, Figure 36B). To further characterize whether some TrBCs were more similar to esoBCs, the two scRNA-seq datasets of TrBCs and esoBCs were combined and analyzed together. Cells from the same tissue origin clustered together, as expected (Figure 44A left panel). General BC markers, Sox2, p63, Krt5 were distributed in both TrBC and esoBC populations. Nkx2-1 specifically marked TrBCs while Pitx1 specifically marked esoBCs (Figure 44A right panels).

5 clusters were identified using DBSCAN. C1 and C2 were from TrBCs; C3, C4 and C5 were from esoBCs (Figure 44B). To quantitatively assess the general features of these five clusters, GSEA was performed to evaluate the esophageal score and tracheal score. Specifically, the esophageal signature and tracheal signature (60-80 genes) were generated by differential gene expression analysis of bulk RNA-seq data of adult TrBCs and esoBCs (N=6 for each). Tracheal signature was defined as gene sets with $\log_2(RPM+1)>5$ and $p<E-5$; while esophageal signature was defined as gene sets with $\log_2(RPM+1)>3$ and $p<E-3$. Consistent with the hypothesis, C2, a subpopulation of TrBCs exhibited slightly higher esophageal score and slightly lower tracheal score comparing to C1, suggesting that C2 might represent a transitional cellular status between TrBCs and esoBCs (Figure 44C). Interestingly, the three esophageal clusters exhibited elevated esophageal scores from C5 to C4 to C3 (Figure 44C left panel). This might indicate that the esophageal score might be a good strategy to assess the differentiating status. Overall, a developing trajectory along C1-C2-C5-C4-C3 could be observed with elevated esophageal features and decreased tracheal features.

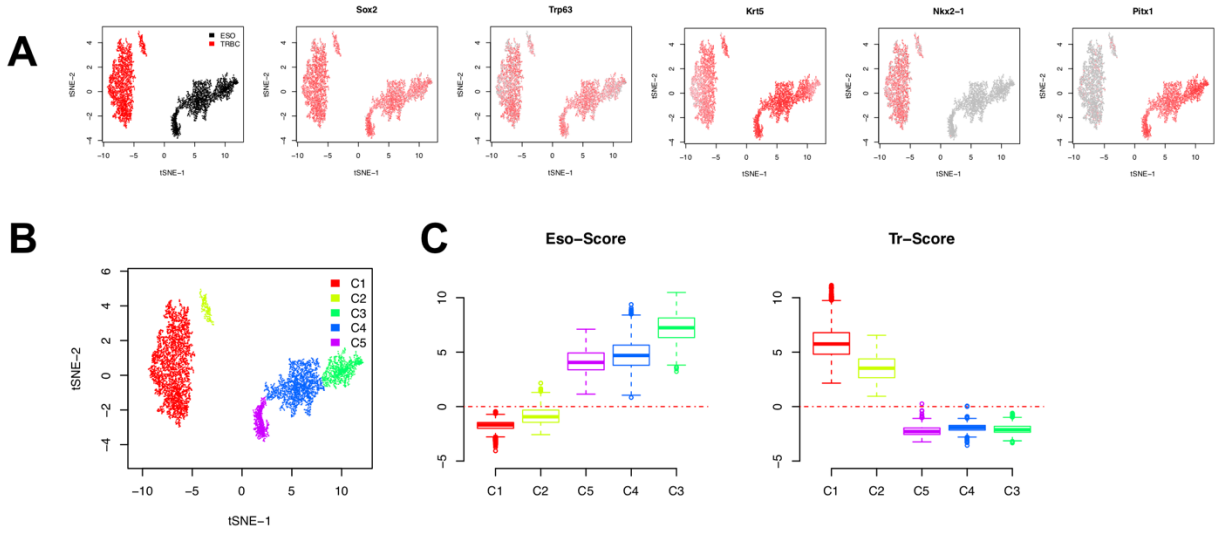


Figure 44. A subpopulation of TrBCs exhibiting esoBC-like features in transcriptome analysis.

(A) t-SNE visualization of scRNA-seq profiles from both TrBCs and esoBCs, colored by cellular origin (left, TrBC in red, esoBC in black), or the expression level ($\log_2(TPM+1)$) of each marker gene. **(B)** 5 distinct clusters (2 in TrBCs, 3 in esoBCs) revealed by density-based spatial clustering of applications with noise

(DBSCAN). **(C)** The distribution of esophageal scores and tracheal scores (generated using the most differentially expressed genes from bulk RNA-seq analysis) for 5 clusters.

To figure out which genes showed continuous dynamics along the observed trajectory, the expression patterns of different genes were characterized and their mean expression levels in each cluster were determined. Two transcription factors, Pitx1 and Barx2 were equally expressed at high levels in 3 esoBC clusters. C2 of TrBC exhibited higher expression of these two genes comparing to C1, suggested a gene regulatory level switch towards an esoBC-like status (Figure 45). Pitx1 (paired like homeodomain 1), was reported to function as regulator for squamous carcinogenesis. In tumor propagating cells, Pitx1 cooperates with Sox2 and p63 to promote self-renewal and to repress Klf4 expression to blunt squamous differentiation[133]. In esophagus and oral epithelium, Pitx1 was specifically expressed in BCs and its reduction had been associated with Barrett's metaplasia, suggesting a functional importance of Pitx1 in regulating squamous differentiation in these epithelia[154]. Barx2 (BARX homeobox 2) was reported as a tumor suppressor gene through inhibiting WNT signaling pathway in NSCLC[155]. Barx2 was expressed in differentiating stratified epithelia from early development, including tongue, nasal epithelium, esophagus, skin[156]. The seemly antagonistic roles of Pitx1 and Barx2 in promoting self-renewal (Pitx1) or squamous differentiation (Barx2) might underlie a tumbledown transcriptional hub that was competent enough to commit differentiation programs of both TrBCs and esoBCs and was vulnerable to perturbation in metaplastic transformation.

Tgfbi and Igfbp2 were highly expressed in three esoBC clusters and exhibited elevated levels in TrBC C2 comparing to C1 (Figure 45). In the bulk RNA-seq analysis for adult TrBCs and esoBCs, Tgfbi was highly expressed in esoBCs but not TrBCs, suggesting that the TGFB/BMP signaling pathway might play essential roles in defining the cellular identities of esoBCs (Figure 34E). Interestingly, TGFB is a major factor in driving squamous metaplasia of airway epithelial cells, and TGFBI ranked no. 10 among the top genes significantly induced during the *in vitro* human bronchial epithelial cells' serial passage (which led to a gradual acquisition of a squamous cell phenotype)[157]. The IGF pathway is important in regulating cell growth, proliferation, metabolism. Although the detailed functions of Igfbp2 in BCs remained to be explored, it had been suggested in the diagnostic marker signature for esophageal squamous cell carcinoma

(ESCC)[158]. The potential activation of TGFB/BMP and IGF pathways in C2 might suggest that squamous metaplasia or malignant transformation may involve with further perturbation of these signaling pathways.

Cytokeratins are intermediate filaments that could serve as markers for epithelial cells in different differentiation statuses or pathological conditions. Krt14 was widely expressed in esoBCs but restricted to a small population of TrBCs. C2 showed significantly higher expression of Krt14, suggesting that these cells differed from the C1 at least in squamous differentiation potential (Figure 45). Watson et al., 2015[4] reported that Krt14 was equally distributed in BSCs and BLPs, and was not a reliable marker of stem cell identity. This was not conflicted with the current observation, since the present data suggested that Krt14 might serve as a marker for cells with more similar features as esoBCs in squamous differentiation potential, not necessarily reflecting the self-renewal capacity. This hypothesis was further corroborated by the observation that squamous differentiation markers Krt13, Dmkn (Dermokine, important in epidermis terminal differentiation[159]), Lgals7 (galectin 7, specifically expressed in keratinocytes and found mainly in stratified squamous epithelium[160]) and others were consistently elevated from C1 to C2 in TrBCs, and from C5/C4 (less differentiated) to C3 (more differentiated) (Figure 45).

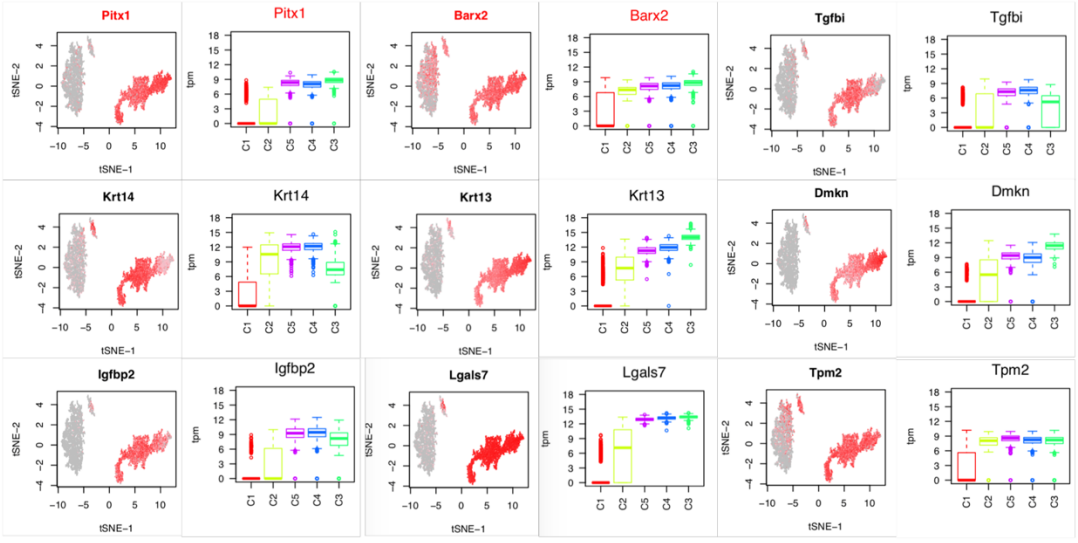


Figure 45. Genes exhibiting increased expression trend along the eso-score trajectory.

t-SNE plot showed scRNA-seq profiles from both TrBCs and esoBCs colored by the expression level (log_2(TPM+1)) of each representative gene on the left. Boxplot showed the distribution of the gene expression level (log_2(TPM+1)). Transcription factors were colored in red.

Many airway-specific genes (Nkx2-1, Epas1, Fos) exhibited reduced expression from C1 to C2, implying an attenuated respiratory cell identity in C2 (Figure 46). Genes related to airway-specific differentiation programs or pathological conditions such as asthma, including Bpifa1, Scgb3a2, Lypd, Sult1d1, Lbp, Aqp4, Sftpd, Slc15a2, Abi3bp, Fmo2/Fmo3, Camk2n1, all showed expression reduction from C1 to C2 and significantly low or no expression in three esoBC clusters. Krt18, paired with Krt8, is a marker for simple columnar epithelial cells. EsoBCs in the stratified epithelium did not express this marker. A slight decrease in Krt18 expression was observed in C2, suggesting a putative switch from the pseudostratified-specific differentiation program towards the stratified-specific program in these cells (Figure 46).

Here, the scRNA-seq datasets of TrBCs and esoBCs were combined, revealing a subpopulation of TrBCs with esoBC-like features at least at the gene expression level. The gene expression analysis and the TrBC-esoBC cell identity trajectory revealed that a considerable number of genes, not only differentiation marker genes, signaling pathway components, but transcription factors as well, showed altered expression patterns in the esoBC-like TrBC subpopulation. This suggested that genome-wide rewired regulatory networks in these cells might endow them the squamous differentiation potential. This unstable transcriptional hub might be easily perturbed or hijacked in chronic injury or carcinogenesis, leading to squamous metaplasia or malignant transformation into squamous cell carcinoma.

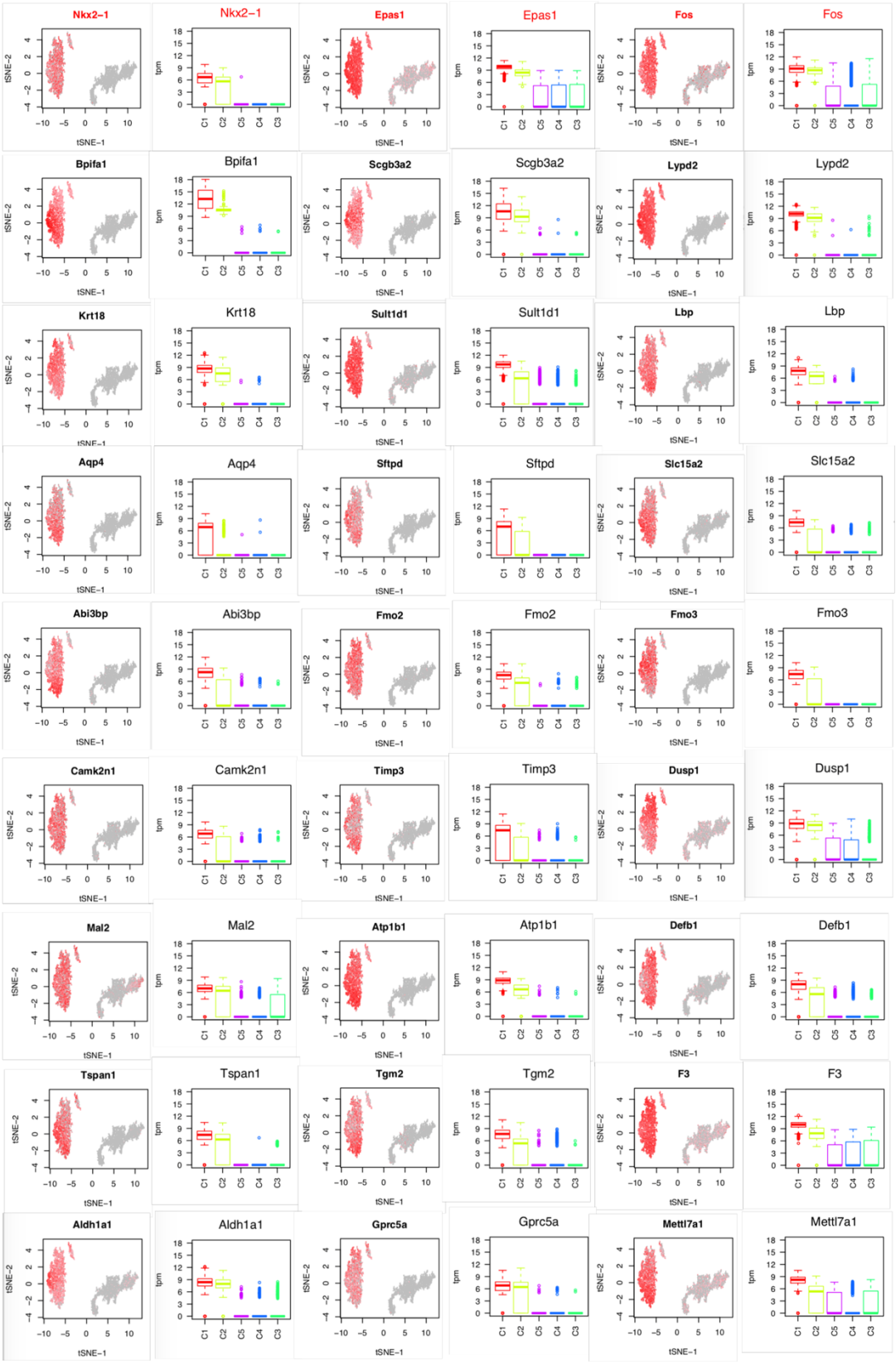


Figure 46. Genes exhibiting decreased expression trend along the tr-score trajectory.

t-SNE plot showed scRNA-seq profiles from both TrBCs and esoBCs colored by the expression level (log_2(TPM+1)) of each representative gene on the left. Boxplot showed the distribution of the gene expression level (log_2(TPM+1)). Transcription factor genes were colored in red.

6-6. Identification of master regulator candidates using metaVIPER analysis

As discussed above, TrBCs and esoBCs represent two adult stem cell populations with mucociliary differentiation program in pseudostratified trachea or squamous differentiation program in stratified esophagus, respectively. They exhibit unexpected cellular plasticity, and could adopt each other's differentiation programs *in vitro* with retinoid acid manipulation or in disease scenarios. The cellular status is represented by a set of genes fulfilling distinct physiological functions, including controlling cell morphology, metabolism, proliferation, differentiation and so on. As gene products interact with each other to form a complex regulatory network (interactome), a handful of transcription factors located at the top of the gene regulatory hierarchy are considered as master regulators for the specific cell status. Perturbation of these few master regulators can collapse the original cellular homeostasis and drive differentiation into another phenotype. This concept has been confirmed in many reprogramming strategies: fibroblasts can be reprogrammed into pluripotent stem cells using Oct4, Sox2, Klf4 and c-Myc[161]; Ascl1, Brn2 and Myt1l can reprogram fibroblasts into functional neurons[162]; overexpression of one single transcription factor Ptf1a can endow a neural stem cell fate in fibroblasts[163]; Pdx1, Ngn3 and MafA can promote the transdifferentiation of pancreatic exocrine cells and liver cells into insulin-secreting pancreatic β cell fate[164-166].

Classic studies used forward genetic screening to identify master regulators controlling cell phenotypes. a handful of transcription factors were selected in the candidate pool based on knowledge and prediction; then a systematic withdrawal or binary search was performed to identify the essential combination of master regulators. This forward strategy relied on previous knowledge and personal understanding, which inevitably introduced human bias. And the determination of final essential master regulator combination was laborious and costly. Recent advances in high-throughput technologies and systems biology solved these problems using a reverse genetic/engineering strategy to predict master regulators from gene expression profiles with high accuracy and specificity. The regulating logic can be reconstructed using

ARACNe algorithm[112]. Based on the ARACNe networks, the expression patterns of transcriptional targets were used as reporters for the protein activities of upstream transcriptional regulators. Transcriptional regulators with higher protein activity scores were considered as candidate master regulators responsible for the gene regulatory network of a specific cellular status[113,167]. Several studies confirmed its robustness and accuracy in inferring master regulators in cancer and tissue development[168-170].

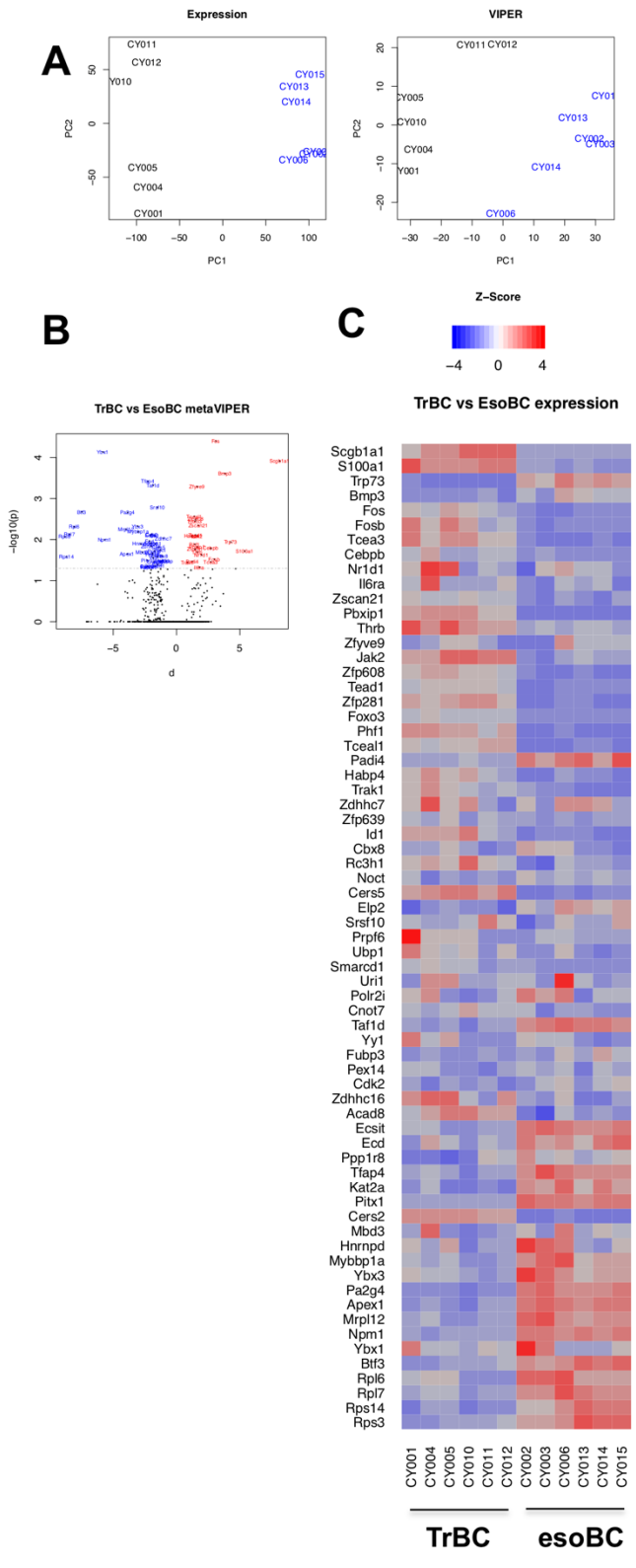
Here, aiming at discovering master regulator candidates which control TrBC and esoBC statuses and perturbation of which leading to cell fate switch, TrBC and esoBC interactomes were built separately using their scRNA-seq datasets. Then these two interactomes were interrogated with bulk RNA-seq transcriptomes of TrBCs and esoBCs through metaVIPER algorithm[114]. At both gene expression level and protein activity level, TrBCs significantly differed from esoBCs (Figure 47A). Through metaVIPER, a list of proteins was identified showing significantly differential activities in either TrBCs (colored in red in Figure 47B) or esoBCs (colored in blue in Figure 47B). As expected, probably due to post-translational regulation, the expression patterns of these candidate genes were not always consistent with their predicted protein activities (Figures 47C & D).

The majority of the candidate master regulators in this list was previously not studied in airway or esophagus. Some genes, such as Fos, Cebpb, Pbxip1, had been identified as COPD-responsive gene and associated with squamous lung cancer[171-175]. Additionally, some genes showing higher protein activities in esoBC status had been suggested to relate with squamous transformation and keratinocyte differentiation. Ybx1 regulates the progenitor functions of epidermal stem cell and is associated with NSCLC[176,177]. Pitx1 plays essential roles in skin squamous cell carcinoma[133]. High expression of Kat2a was found associated with ESCC risk[178]. Id1 has been suggested to promote tumor growth and progression in oral squamous cell carcinoma and NSCLC[179,180]. Interestingly, a lot of epigenetic modifiers (Kat2a, Yy1, Mbd3, Mybbp1a, Phf1) were presented in this list, suggesting a physiological importance of epigenetic regulation in controlling the cellular statuses between TrBCs and esoBCs.

Alternative analysis could be performed as following: instead of building two ARACNe interactomes separately for TrBCs and esoBCs, these two scRNA-seq datasets could be combined and used for one single interactome construction. In this case, more than 8000 scRNA-seq profiles could be used to assess the regulatory logic for general BC status, which will significantly increase the accuracy. Then, the bulk

RNA-seq datasets for both esoBCs and TrBCs could be analyzed through VIPER algorithm, as the esoBCs and TrBCs are probably similar in general epigenetics. It would be interesting to compare the master regulator candidate lists generated from two separated interactomes with meta-VIPER to one combined interactome with VIPER.

Future validation for these candidate master regulators is required. The *in vitro* differentiation assay using air-liquid interface culture and retinoic acid-deficient squamous metaplasia model might represent a suitable platform to test the phenotypic alteration following genetic manipulation. This functional test is phenotype-driven, and highly depends on the reliability of the disease model. Another way to assess the functional importance of candidate master regulators is to use transcriptome as readout instead of phenotype, which should be unbiased but suffers from indirect correlation with final phenotypic outcome. High-throughput methods such as Perturb-seq or CROP-seq[181,182], combining CRISPR-based gene expression perturbation and scRNA-seq, can be used to evaluate the importance of candidate master regulators in controlling gene regulatory networks. A combination of both assays will facilitate the functional validation of the candidate pool and provide mechanistic views to explain cell fate regulation from distinct angles.



D

	d	p
Scgb1a1	8.141	0.000
S100a1	5.382	0.019
Trp73	4.283	0.012
Bmp3	3.829	0.000
Fos	3.105	0.000
Fosb	2.983	0.030
Tcea3	2.710	0.035
Cebpb	2.707	0.016
Nr1d1	1.968	0.023
Il6ra	1.811	0.047
Zscan21	1.727	0.005
Pbxip1	1.640	0.015
Thrb	1.640	0.020
Zfyve9	1.632	0.001
Jak2	1.629	0.008
Zfp608	1.516	0.003
Tead1	1.501	0.008
Zfp281	1.482	0.017
Foxo3	1.479	0.004
Phf1	1.443	0.013
Tceal1	1.409	0.003
Padi4	1.249	0.034
Habp4	1.207	0.008
Trak1	0.869	0.036
Zdhhc7	-1.027	0.009
Zfp639	-1.143	0.033
Id1	-1.147	0.032
Cbx8	-1.154	0.025
Rc3h1	-1.248	0.035
Noct	-1.282	0.018
Cers5	-1.411	0.015
Elp2	-1.498	0.040
Srsf10	-1.504	0.002
Prpf6	-1.519	0.026
Ubp1	-1.682	0.013
Smarcd1	-1.715	0.044
Uri1	-1.723	0.008
Polr2i	-1.745	0.019
Cnot7	-1.789	0.035
Taf1d	-1.853	0.000
Yy1	-1.853	0.028
Fubp3	-1.896	0.022
Pex14	-1.898	0.023
Cdk2	-1.969	0.007
Zdhhc16	-2.013	0.014
Acad8	-2.031	0.012
Ecsit	-2.057	0.008
Ecd	-2.143	0.011
Ppp1r8	-2.178	0.048
Tfap4	-2.266	0.000
Kat2a	-2.303	0.046
Pitx1	-2.326	0.032
Cers2	-2.329	0.044
Mbd3	-2.704	0.020
Hnrnpd	-2.819	0.013
Mybbp1a	-3.025	0.006
Ybx3	-3.068	0.005
Pa2g4	-3.859	0.002
Apex1	-3.894	0.022
Mrpl12	-3.984	0.006
Npm1	-5.672	0.010
Ybx1	-5.821	0.000
Btf3	-7.488	0.002
Rpl6	-8.081	0.005
Rpl7	-8.440	0.007
Rps14	-8.690	0.026
Rps3	-8.818	0.009

Figure 47. Candidate master regulators inferred by metaVIPER analysis.

(A) Principal component analysis (PCA) showing distinct difference between bulk RNA-seq profiles (left panel) and protein activity profiles (right panel) of adult TrBCs (black) and esoBCs (blue). **(B)** Volcano plot showing distinct protein activities in TrBCs and esoBCs. Differential protein activity scores for all genes were represented in black dots; genes showing significantly higher protein activities in TrBCs were labeled with gene names in red ($p<0.05$; $d>0$); genes showing significantly higher protein activities in esoBCs were labeled with gene names in blue ($p<0.05$; $d<0$). **(C)** Relative expression levels of genes (row-wise Z-score of gene expression) inferred as candidate master regulators based on significantly distinct protein activities in TrBCs and esoBCs. **(D)** Table showing candidate master regulators with differential protein activity scores (d) and corresponding t-test p values (p).

Chapter 7. Conclusion

This thesis shows that when airways are still forming, two lineage-restriction events occur in p63-expressing cells, which ultimately establish regionally distinct adult multipotent progenitor pools: the BCs in trachea and the $p63^+$ $Krt5^-$ cells in intrapulmonary airways (Figure 48).

In this study, an early onset of p63 expression was identified during specification of the lung field in anterior foregut endoderm since E9.0. Unexpectedly, these early $p63^+$ cells are multipotent to generate all cell types of both airways and alveoli. At this time, the proximal p63 lineage has not been separated from the distal Sox9/Id2 lineage. From E10.5, a first proximal-distal lineage segregation event occurs. $p63^+$ cells permanently lose the potential for alveolar differentiation, and become restricted to extrapulmonary airways. $Sox9^+$ $Id2^+$ distal epithelial progenitors mainly contribute to form the cell types populating intrapulmonary airways and alveolar compartment. A second basal-luminal lineage segregation event occurs at E13.5 in the trachea. Some of the $p63^+$ cells labeled in the *p63-CreERT2* lineage tracing system (potentially with higher p63 promoter activity) acquire BC fate and self-renew to expand the BC pool, while others commit to luminal differentiation directly.

A population of embryonic $p63^+$ cells is maintained immature and uncommitted in intrapulmonary large airways throughout lifetime. In adulthood upon H1N1 challenge, these $p63^+$ $Krt5^-$ cells are recruited support the injured lung. They expand and migrate along the proximal airways towards the area of severe alveolar damage to form the $Krt5^+$ pods. The specification/maintenance of intrapulmonary $p63^+$ $Krt5^-$ progenitor pool depends on p63 gene dosage. Loss of one functional allele of p63 gene leads to a less effective response of this progenitor pool and ultimately a more attenuated response. The present study shows minimal if any alveolar regeneration from the ectopic $Krt5^+$ pods, further suggesting that the alveolar differentiation potential of early multipotent $p63^+$ progenitors is lost as early as E10.5, during the first lineage-segregation event. It is of great scientific interests and clinical value to investigate how this lineage restriction occurs and how one can stimulate the alveolar differentiation potential in these adult $Krt5^+$ pods post injury.

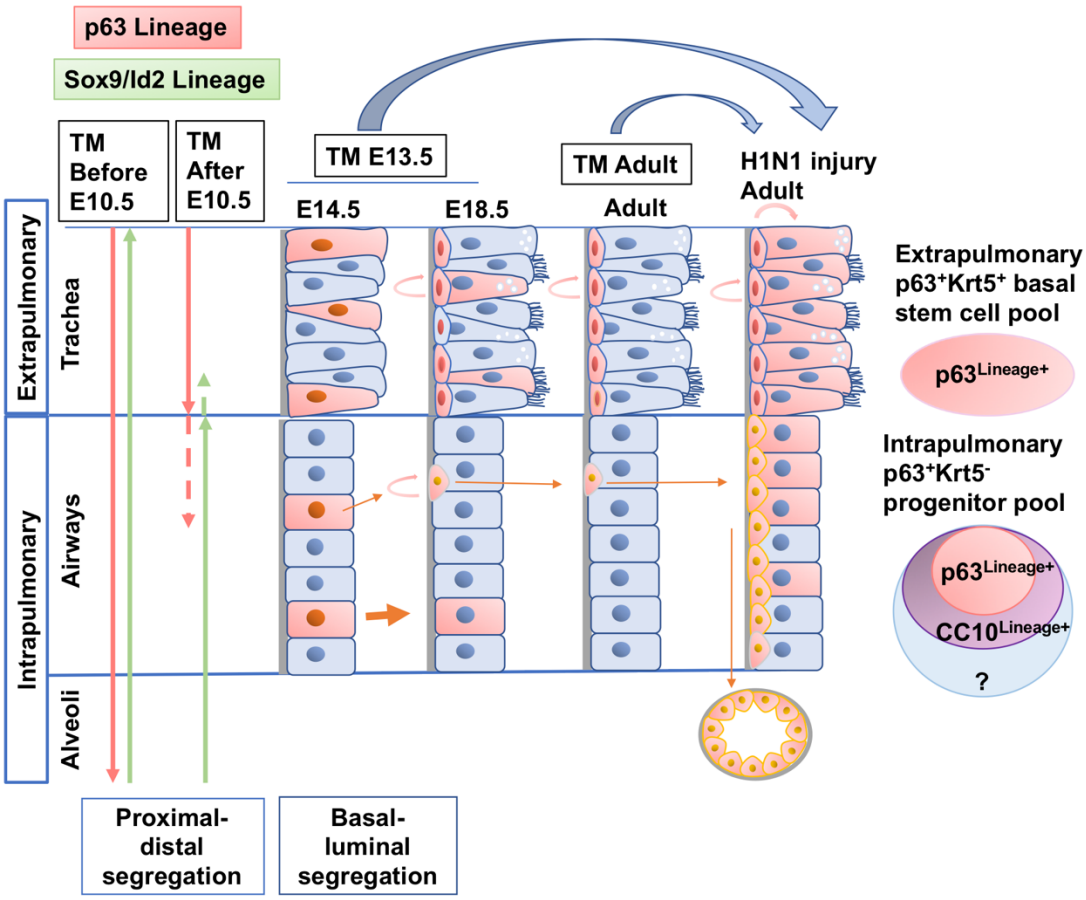


Figure 48. Schematic presentation showing the contribution of p63 lineage during airway development and injury response.

In the last part of this thesis, molecular signatures are generated for E18.5 preBCs in trachea, adult TrBCs and esoBCs. This reveals several signaling pathways which might play essential roles in TrBC maturation or the fate maintenance of adult TrBCs/esoBCs. The scRNA-seq datasets generated here uncover underappreciated heterogeneity in adult stem cell populations in trachea and esophagus. TrBCs contain clusters of cells with different levels of squamous differentiation potential, which may underlie the significant plasticity exhibited in chronic injury response or pathological metaplasia. Furthermore, candidate master regulators were identified controlling distinct cellular statuses using both expression analysis and protein activity analysis, which await further functional characterization and validation.

Taken together, this thesis research provides insights into the origin and diversity of the adult stem cells in the respiratory tract, and the lineage relationships arising from responses to environmental agents.

References

1. Rock JR, Randell SH, Hogan BL. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis Model Mech.* 2010;3(9-10):545-556.

2. Rock JR, Onaitis MW, Rawlins EL, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A.* 2009;106(31):12771-12775.

3. Rock JR, Gao X, Xue Y, Randell SH, Kong YY, Hogan BL. Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell.* 2011;8(6):639-648.

4. Watson JK, Rulands S, Wilkinson AC, et al. Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. *Cell Rep.* 2015;12(1):90-101.

5. Melino G, Memmi EM, Pelicci PG, Bernassola F. Maintaining epithelial stemness with p63. *Sci Signal.* 2015;8(387):re9.

6. Ghosh M, Brechbuhl HM, Smith RW, et al. Context-dependent differentiation of multipotential keratin 14-expressing tracheal basal cells. *Am J Respir Cell Mol Biol.* 2011;45(2):403-410.

7. Hogan BL, Barkauskas CE, Chapman HA, et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell.* 2014;15(2):123-138.

8. Rock JR, Hogan BL. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu Rev Cell Dev Biol.* 2011;27:493-512.

9. Bilodeau M, Shojaie S, Ackerley C, Post M, Rossant J. Identification of a proximal progenitor population from murine fetal lungs with clonogenic and multilineage differentiation potential. *Stem Cell Reports.* 2014;3(4):634-649.

10. Que J, Okubo T, Goldenring JR, et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development.* 2007;134(13):2521-2531.

11. Rackley CR, Stripp BR. Building and maintaining the epithelium of the lung. *J Clin Invest.* 2012;122(8):2724-2730.

12. Burgel PR, Bergeron A, de Blic J, et al. Small airways diseases, excluding asthma and COPD: an overview. *Eur Respir Rev.* 2013;22(128):131-147.

13. Montoro DT, Haber AL, Biton M, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature.* 2018;560(7718):319-324.

14. Rawlins EL, Ostrowski LE, Randell SH, Hogan BL. Lung development and repair: contribution of the ciliated lineage. *Proc Natl Acad Sci U S A.* 2007;104(2):410-417.

15. Firth AL, Dargitz CT, Qualls SJ, et al. Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proc Natl Acad Sci U S A.* 2014;111(17):E1723-1730.

16. Guha A, Vasconcelos M, Cai Y, et al. Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. *Proc Natl Acad Sci U S A.* 2012;109(31):12592-12597.

17. Guha A, Vasconcelos M, Zhao R, Gower AC, Rajagopal J, Cardoso WV. Analysis of Notch signaling-dependent gene expression in developing airways reveals diversity of Clara cells. *PLoS One.* 2014;9(2):e88848.

18. McCauley KB, Alysandratos KD, Jacob A, et al. Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium. *Stem Cell Reports.* 2018;10(5):1579-1595.

19. Branchfield K, Nantie L, Verheyden JM, Sui P, Wienhold MD, Sun X. Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. *Science.* 2016;351(6274):707-710.

20. Sui P, Wiesner DL, Xu J, et al. Pulmonary neuroendocrine cells amplify allergic asthma responses. *Science.* 2018;360(6393).

21. Yang J, Hernandez BJ, Martinez Alanis D, et al. The development and plasticity of alveolar type 1 cells. *Development.* 2016;143(1):54-65.

22. Que J, Choi M, Ziel JW, Klingensmith J, Hogan BL. Morphogenesis of the trachea and esophagus: current players and new roles for noggin and Bmps. *Differentiation.* 2006;74(7):422-437.

23. Domyan ET, Ferretti E, Throckmorton K, Mishina Y, Nicolis SK, Sun X. Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development.* 2011;138(5):971-981.

24. Serls AE, Doherty S, Parvatiyar P, Wells JM, Deutsch GH. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development.* 2005;132(1):35-47.

25. Goss AM, Tian Y, Tsukiyama T, et al. Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev Cell.* 2009;17(2):290-298.

26. Harris-Johnson KS, Domyan ET, Vezina CM, Sun X. beta-Catenin promotes respiratory progenitor identity in mouse foregut. *Proc Natl Acad Sci U S A.* 2009;106(38):16287-16292.

27. Motoyama J, Liu J, Mo R, Ding Q, Post M, Hui CC. Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat Genet.* 1998;20(1):54-57.

28. Chen F, Desai TJ, Qian J, Niederreither K, Lu J, Cardoso WV. Inhibition of Tgf beta signaling by endogenous retinoic acid is essential for primary lung bud induction. *Development.* 2007;134(16):2969-2979.

29. Rankin SA, Han L, McCracken KW, et al. A Retinoic Acid-Hedgehog Cascade Coordinates Mesoderm-Inducing Signals and Endoderm Competence during Lung Specification. *Cell Rep.* 2016;16(1):66-78.

30. Snyder EL, Watanabe H, Magendantz M, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell.* 2013;50(2):185-199.

31. Que J, Luo X, Schwartz RJ, Hogan BL. Multiple roles for Sox2 in the developing and adult mouse trachea. *Development.* 2009;136(11):1899-1907.

32. Jacobs IJ, Ku WY, Que J. Genetic and cellular mechanisms regulating anterior foregut and esophageal development. *Dev Biol.* 2012;369(1):54-64.

33. Que J. The initial establishment and epithelial morphogenesis of the esophagus: a new model of tracheal-esophageal separation and transition of simple columnar into stratified squamous epithelium in the developing esophagus. *Wiley Interdiscip Rev Dev Biol.* 2015;4(4):419-430.

34. Sasaki T, Kusafuka T, Okada A. Analysis of the development of normal foregut and tracheoesophageal fistula in an adriamycin rat model using three-dimensional image reconstruction. *Surg Today.* 2001;31(2):133-139.

35. Zhang Y, Jiang M, Kim E, et al. Development and stem cells of the esophagus. *Semin Cell Dev Biol.* 2017;66:25-35.

36. Arman E, Haffner-Krausz R, Gorivodsky M, Lonai P. Fgfr2 is required for limb outgrowth and lung-branching morphogenesis. *Proc Natl Acad Sci U S A.* 1999;96(21):11895-11899.

37. De Moerlooze L, Spencer-Dene B, Revest JM, Hajihosseini M, Rosewell I, Dickson C. An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal-epithelial signalling during mouse organogenesis. *Development.* 2000;127(3):483-492.

38. Sekine K, Ohuchi H, Fujiwara M, et al. Fgf10 is essential for limb and lung formation. *Nat Genet.* 1999;21(1):138-141.

39. Cardoso WV, Lu J. Regulation of early lung morphogenesis: questions, facts and controversies. *Development.* 2006;133(9):1611-1624.

40. Rawlins EL, Clark CP, Xue Y, Hogan BL. The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development.* 2009;136(22):3741-3745.

41. Alanis DM, Chang DR, Akiyama H, Krasnow MA, Chen J. Two nested developmental waves demarcate a compartment boundary in the mouse lung. *Nat Commun.* 2014;5:3923.

42. Weaver M, Yingling JM, Dunn NR, Bellusci S, Hogan BL. Bmp signaling regulates proximal-distal differentiation of endoderm in mouse lung development. *Development.* 1999;126(18):4005-4015.

43. Ostrin EJ, Little DR, Gerner-Mauro KN, et al. beta-Catenin maintains lung epithelial progenitors after lung specification. *Development.* 2018;145(5).

44. Mahoney JE, Mori M, Szymaniak AD, Varelas X, Cardoso WV. The hippo pathway effector Yap controls patterning and differentiation of airway epithelial progenitors. *Dev Cell.* 2014;30(2):137-150.

45. Szymaniak AD, Mahoney JE, Cardoso WV, Varelas X. Crumbs3-Mediated Polarity Directs Airway Epithelial Cell Fate through the Hippo Pathway Effector Yap. *Dev Cell.* 2015;34(3):283-296.

46. Tsao PN, Vasconcelos M, Izvolsky KI, Qian J, Lu J, Cardoso WV. Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development.* 2009;136(13):2297-2307.

47. Guseh JS, Bores SA, Stanger BZ, et al. Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. *Development.* 2009;136(10):1751-1759.

48. Morimoto M, Nishinakamura R, Saga Y, Kopan R. Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development.* 2012;139(23):4365-4373.

49. Treutlein B, Brownfield DG, Wu AR, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature.* 2014;509(7500):371-375.

50. Hong KU, Reynolds SD, Watkins S, Fuchs E, Stripp BR. Basal cells are a multipotent progenitor capable of renewing the bronchial epithelium. *Am J Pathol.* 2004;164(2):577-588.

51. Fulcher ML, Gabriel S, Burns KA, Yankaskas JR, Randell SH. Well-differentiated human airway epithelial cell cultures. *Methods Mol Med.* 2005;107:183-206.

52. Liu JY, Nettesheim P, Randell SH. Growth and differentiation of tracheal epithelial progenitor cells. *Am J Physiol.* 1994;266(3 Pt 1):L296-307.

53. Gao X, Bali AS, Randell SH, Hogan BL. GRHL2 coordinates regeneration of a polarized mucociliary epithelium from basal stem cells. *J Cell Biol.* 2015;211(3):669-682.

54. Pardo-Saganta A, Law BM, Tata PR, et al. Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell.* 2015;16(2):184-197.

55. Tadokoro T, Wang Y, Barak LS, Bai Y, Randell SH, Hogan BL. IL-6/STAT3 promotes regeneration of airway ciliated cells from basal stem cells. *Proc Natl Acad Sci U S A.* 2014;111(35):E3641-3649.

56. Lu L, Teixeira VH, Yuan Z, et al. LRIG1 regulates cadherin-dependent contact inhibition directing epithelial homeostasis and pre-invasive squamous cell carcinoma development. *J Pathol.* 2013;229(4):608-620.

57. Brechbuhl HM, Li B, Smith RW, Reynolds SD. Epidermal growth factor receptor activity is necessary for mouse basal cell proliferation. *Am J Physiol Lung Cell Mol Physiol.* 2014;307(10):L800-810.

58. Giangreco A, Lu L, Vickers C, et al. beta-Catenin determines upper airway progenitor cell fate and preinvasive squamous lung cancer progression by modulating epithelial-mesenchymal transition. *J Pathol.* 2012;226(4):575-587.

59. Zhao R, Fallon TR, Saladi SV, et al. Yap tunes airway epithelial size and architecture by regulating the identity, maintenance, and self-renewal of stem cells. *Dev Cell.* 2014;30(2):151-165.

60. Balasooriya GI, Johnson JA, Basson MA, Rawlins EL. An FGFR1-SPRY2 Signaling Axis Limits Basal Cell Proliferation in the Steady-State Airway Epithelium. *Dev Cell.* 2016;37(1):85-97.

61. Balasooriya GI, Goschorska M, Piddini E, Rawlins EL. FGFR2 is required for airway basal cell self-renewal and terminal differentiation. *Development.* 2017;144(9):1600-1606.

62. Mou H, Vinarsky V, Tata PR, et al. Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. *Cell Stem Cell.* 2016;19(2):217-231.

63. Volckaert T, Campbell A, Dill E, Li C, Minoo P, De Langhe S. Localized Fgf10 expression is not required for lung branching morphogenesis but prevents differentiation of epithelial progenitors. *Development.* 2013;140(18):3731-3742.

64. Snitow ME, Li S, Morley MP, et al. Ezh2 represses the basal cell lineage during lung endoderm development. *Development.* 2015;142(1):108-117.

65. Galvis LA, Holik AZ, Short KM, et al. Repression of Igf1 expression by Ezh2 prevents basal cell differentiation in the developing lung. *Development.* 2015;142(8):1458-1469.

66. Crum CP, McKeon FD. p63 in epithelial survival, germ cell surveillance, and neoplasia. *Annu Rev Pathol.* 2010;5:349-371.

67. Fletcher RB, Prasol MS, Estrada J, et al. p63 regulates olfactory stem cell self-renewal and differentiation. *Neuron.* 2011;72(5):748-759.

68. Yang A, Schweitzer R, Sun D, et al. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature.* 1999;398(6729):714-718.

69. Mills AA, Zheng B, Wang XJ, Vogel H, Roop DR, Bradley A. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature.* 1999;398(6729):708-713.

70. Vanbokhoven H, Melino G, Candi E, Declercq W. p63, a story of mice and men. *J Invest Dermatol.* 2011;131(6):1196-1207.

71. Senoo M, Pinto F, Crum CP, McKeon F. p63 Is essential for the proliferative potential of stem cells in stratified epithelia. *Cell.* 2007;129(3):523-536.

72. Su X, Paris M, Gi YJ, et al. TAp63 prevents premature aging by promoting adult stem cell maintenance. *Cell Stem Cell.* 2009;5(1):64-75.

73. Nemajerova A, Kramer D, Siller SS, et al. TAp73 is a central transcriptional regulator of airway multiciliogenesis. *Genes Dev.* 2016;30(11):1300-1312.

74. Rouleau M, Medawar A, Hamon L, et al. TAp63 is important for cardiac differentiation of embryonic stem cells and heart development. *Stem Cells.* 2011;29(11):1672-1683.

75. Wang Q, Zou Y, Nowotschin S, et al. The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell.* 2017;20(1):70-86.

76. Pignon JC, Grisanzio C, Geng Y, Song J, Shivdasani RA, Signoretti S. p63-expressing cells are the stem cells of developing prostate, bladder, and colorectal epithelia. *Proc Natl Acad Sci U S A.* 2013;110(20):8105-8110.

77. Romano RA, Smalley K, Magraw C, et al. DeltaNp63 knockout mice reveal its indispensable role as a master regulator of epithelial development and differentiation. *Development.* 2012;139(4):772-782.

78. Koster MI, Kim S, Mills AA, DeMayo FJ, Roop DR. p63 is the molecular switch for initiation of an epithelial stratification program. *Genes Dev.* 2004;18(2):126-131.

79. Li L, Wang Y, Torkelson JL, et al. TFAP2C- and p63-Dependent Networks Sequentially Rearrange Chromatin Landscapes to Drive Human Epidermal Lineage Commitment. *Cell Stem Cell.* 2019;24(2):271-284 e278.

80. Pattison JM, Melo SP, Piekos SN, et al. Retinoic acid and BMP4 cooperate with p63 to alter chromatin dynamics during surface epithelial commitment. *Nat Genet.* 2018;50(12):1658-1665.

81. Chakrabarti R, Wei Y, Hwang J, et al. DeltaNp63 promotes stem cell activity in mammary gland development and basal-like breast cancer by enhancing Fzd7 expression and Wnt signalling. *Nat Cell Biol.* 2014;16(10):1004-1015, 1001-1013.

82. Daniely Y, Liao G, Dixon D, et al. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. *Am J Physiol Cell Physiol.* 2004;287(1):C171-181.

83. Arason AJ, Jonsdottir HR, Halldorsson S, et al. deltaNp63 has a role in maintaining epithelial integrity in airway epithelium. *PLoS One.* 2014;9(2):e88683.

84. Rawlins EL, Okubo T, Xue Y, et al. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell.* 2009;4(6):525-534.

85. Guha A, Deshpande A, Jain A, Sebastiani P, Cardoso WV. Uroplakin 3a(+) Cells Are a Distinctive Population of Epithelial Progenitors that Contribute to Airway Maintenance and Post-injury Repair. *Cell Rep.* 2017;19(2):246-254.

86. Barkauskas CE, Cronce MJ, Rackley CR, et al. Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest.* 2013;123(7):3025-3036.

87. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature.* 2014;507(7491):190-194.

88. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science.* 2018;359(6380):1118-1123.

89. Zacharias WJ, Frank DB, Zepp JA, et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature.* 2018;555(7695):251-255.

90. Tata PR, Mou H, Pardo-Saganta A, et al. Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature.* 2013;503(7475):218-223.

91. Jain R, Barkauskas CE, Takeda N, et al. Plasticity of Hopx(+) type I alveolar cells to regenerate type II cells in the lung. *Nat Commun.* 2015;6:6727.

92. Song H, Yao E, Lin C, Gacayan R, Chen MH, Chuang PT. Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. *Proc Natl Acad Sci U S A.* 2012;109(43):17531-17536.

93. Snippert HJ, van der Flier LG, Sato T, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell.* 2010;143(1):134-144.

94. Yan KS, Chia LA, Li X, et al. The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc Natl Acad Sci U S A.* 2012;109(2):466-471.

95. Yu S, Tong K, Zhao Y, et al. Paneth Cell Multipotency Induced by Notch Activation following Injury. *Cell Stem Cell.* 2018;23(1):46-59 e45.

96. Tata A, Kobayashi Y, Chow RD, et al. Myoepithelial Cells of Submucosal Glands Can Function as Reserve Stem Cells to Regenerate Airways after Injury. *Cell Stem Cell.* 2018;22(5):668-683 e666.

97. Lynch TJ, Anderson PJ, Rotti PG, et al. Submucosal Gland Myoepithelial Cells Are Reserve Stem Cells That Can Regenerate Mouse Tracheal Epithelium. *Cell Stem Cell.* 2018;22(5):653-667 e655.

98. Yang Y, Cardoso WV. Stem Cells Sheltered from Air-Raids Repair Airways. *Cell Stem Cell.* 2018;22(5):613-614.

99. Bonfanti P, Claudinot S, Amici AW, Farley A, Blackburn CC, Barrandon Y. Microenvironmental reprogramming of thymic epithelial cells to skin multipotent stem cells. *Nature.* 2010;466(7309):978-982.

100. Rompolas P, Mesa KR, Greco V. Spatial organization within a niche as a determinant of stem-cell fate. *Nature.* 2013;502(7472):513-518.

101. Kumar PA, Hu Y, Yamamoto Y, et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell.* 2011;147(3):525-538.

102. Zuo W, Zhang T, Wu DZ, et al. p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration. *Nature.* 2015;517(7536):616-620.

103. Zheng D, Yin L, Chen J. Evidence for Scgb1a1(+) cells in the generation of p63(+) cells in the damaged lung parenchyma. *Am J Respir Cell Mol Biol.* 2014;50(3):595-604.

104. Zheng D, Limmon GV, Yin L, et al. Regeneration of alveolar type I and II cells from Scgb1a1-expressing cells following severe pulmonary damage induced by bleomycin and influenza. *PLoS One.* 2012;7(10):e48451.

105. Vaughan AE, Brumwell AN, Xi Y, et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature.* 2015;517(7536):621-625.

106. Ray S, Chiba N, Yao C, et al. Rare SOX2(+) Airway Progenitor Cells Generate KRT5(+) Cells that Repopulate Damaged Alveolar Parenchyma following Influenza Virus Infection. *Stem Cell Reports.* 2016;7(5):817-825.

107. Xi Y, Kim T, Brumwell AN, et al. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nat Cell Biol.* 2017;19(8):904-914.

108. Lee DK, Liu Y, Liao L, Wang F, Xu J. The prostate basal cell (BC) heterogeneity and the p63-positive BC differentiation spectrum in mice. *Int J Biol Sci.* 2014;10(9):1007-1017.

109. Soeda T, Deng JM, de Crombrugghe B, Behringer RR, Nakamura T, Akiyama H. Sox9-expressing precursors are the cellular origin of the cruciate ligament of the knee joint and the limb tendons. *Genesis.* 2010;48(11):635-644.

110. Fre S, Hannezo E, Sale S, et al. Notch lineages and activity in intestinal stem cells determined by a new set of knock-in mice. *PLoS One.* 2011;6(10):e25785.

111. Huang SX, Green MD, de Carvalho AT, et al. The in vitro generation of lung and airway progenitor cells from human pluripotent stem cells. *Nat Protoc.* 2015;10(3):413-425.

112. Margolin AA, Nemenman I, Basso K, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics.* 2006;7 Suppl 1:S7.

113. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet.* 2016;48(8):838-847.

114. Ding H, Douglass EF, Jr., Sonabend AM, et al. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun.* 2018;9(1):1471.

115. Ding H, Wang W, Califano A. iterClust: a statistical framework for iterative clustering analysis. *Bioinformatics.* 2018;34(16):2865-2866.

116. Anderson PJ, Lynch TJ, Engelhardt JF. Multipotent Myoepithelial Progenitor Cells Are Born Early during Airway Submucosal Gland Development. *Am J Respir Cell Mol Biol.* 2017;56(6):716-726.

117. Xie W, Lynch TJ, Liu X, et al. Sox2 modulates Lef-1 expression during airway submucosal gland development. *Am J Physiol Lung Cell Mol Physiol.* 2014;306(7):L645-660.

118. Marshall CB, Mays DJ, Beeler JS, et al. p73 Is Required for Multiciliogenesis and Regulates the Foxj1-Associated Gene Network. *Cell Rep.* 2016;14(10):2289-2300.

119. Koster MI, Kim S, Roop DR. P63 deficiency: a failure of lineage commitment or stem cell maintenance? *J Investig Dermatol Symp Proc.* 2005;10(2):118-123.

120. McKeon F. p63 and the epithelial stem cell: more than status quo? *Genes Dev.* 2004;18(5):465-469.

121. Xing Y, Li C, Li A, et al. Signaling via Alk5 controls the ontogeny of lung Clara cells. *Development.* 2010;137(5):825-833.

122. Mori M, Mahoney JE, Stupnikov MR, et al. Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development.* 2015;142(2):258-267.

123. Javelaud D, Alexaki VI, Dennler S, Mohammad KS, Guise TA, Mauviel A. TGF-beta/SMAD/GLI2 signaling axis in cancer progression and metastasis. *Cancer Res.* 2011;71(17):5606-5610.

124. Huang L, Walter V, Hayes DN, Onaitis M. Hedgehog-GLI signaling inhibition suppresses tumor growth in squamous lung cancer. *Clin Cancer Res.* 2014;20(6):1566-1575.

125. Regl G, Neill GW, Eichberger T, et al. Human GLI2 and GLI1 are part of a positive feedback mechanism in Basal Cell Carcinoma. *Oncogene.* 2002;21(36):5529-5539.

126. Tang YA, Chen YF, Bao Y, et al. Hypoxic tumor microenvironment activates GLI2 via HIF-1alpha and TGF-beta2 to promote chemoresistance in colorectal cancer. *Proc Natl Acad Sci U S A.* 2018;115(26):E5990-E5999.

127. Tadokoro T, Gao X, Hong CC, Hotten D, Hogan BL. BMP signaling and cellular dynamics during regeneration of airway epithelium from basal progenitors. *Development.* 2016;143(5):764-773.

128. Waldner MJ, Foersch S, Neurath MF. Interleukin-6--a key regulator of colorectal cancer development. *Int J Biol Sci.* 2012;8(9):1248-1253.

129. Tello D, Balsa E, Acosta-Iborra B, et al. Induction of the mitochondrial NDUFA4L2 protein by HIF-1alpha decreases oxygen consumption by inhibiting Complex I activity. *Cell Metab.* 2011;14(6):768-779.

130. Yuen HF, McCrudden CM, Chan KK, et al. The role of Pea3 group transcription factors in esophageal squamous cell carcinoma. *Am J Pathol.* 2011;179(2):992-1003.

131. Watanabe H, Ma Q, Peng S, et al. SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. *J Clin Invest.* 2014;124(4):1636-1645.

132. Lee JH, Tammela T, Hofree M, et al. Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6. *Cell.* 2017;170(6):1149-1163 e1112.

133. Sastre-Perona A, Hoang-Phou S, Leitner MC, Okuniewska M, Meehan S, Schober M. De Novo PITX1 Expression Controls Bi-Stable Transcriptional Circuits to Govern Self-Renewal and Differentiation in Squamous Cell Carcinoma. *Cell Stem Cell.* 2019;24(3):390-404 e398.

134. Eichberger T, Regl G, Ikram MS, et al. FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. *J Invest Dermatol.* 2004;122(5):1180-1187.

135. Kimura-Yoshida C, Mochida K, Nakaya MA, Mizutani T, Matsuo I. Cytoplasmic localization of GRHL3 upon epidermal differentiation triggers cell shape change for epithelial morphogenesis. *Nat Commun.* 2018;9(1):4059.

136. Lee B, Villarreal-Ponce A, Fallahi M, et al. Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells. *Dev Cell.* 2014;29(1):47-58.

137. Sen GL, Boxer LD, Webster DE, et al. ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. *Dev Cell.* 2012;22(3):669-677.

138. Kim W, Kim E, Lee S, et al. TFAP2C-mediated upregulation of TGFBR1 promotes lung tumorigenesis and epithelial-mesenchymal transition. *Exp Mol Med.* 2016;48(11):e273.

139. Yang Y, Riccio P, Schotsaert M, et al. Spatial-Temporal Lineage Restrictions of Embryonic p63(+) Progenitors Establish Distinct Stem Cell Pools in Adult Airways. *Dev Cell.* 2018;44(6):752-761 e754.

140. Cole BB, Smith RW, Jenkins KM, Graham BB, Reynolds PR, Reynolds SD. Tracheal Basal cells: a facultative progenitor cell pool. *Am J Pathol.* 2010;177(1):362-376.

141. Li Y, Xu Y, Ye K, et al. Knockdown of Tubulin Polymerization Promoting Protein Family Member 3 Suppresses Proliferation and Induces Apoptosis in Non-Small-Cell Lung Cancer. *J Cancer.* 2016;7(10):1189-1196.

142. Sunaga N, Miyajima K, Suzuki M, et al. Different roles for caveolin-1 in the development of non-small cell lung cancer versus small cell lung cancer. *Cancer Res.* 2004;64(12):4277-4285.

143. Peng T, Tian Y, Boogerd CJ, et al. Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. *Nature.* 2013;500(7464):589-592.

144. Hines EA, Jones MK, Verheyden JM, Harvey JF, Sun X. Establishment of smooth muscle and cartilage juxtaposition in the developing mouse upper airways. *Proc Natl Acad Sci U S A.* 2013;110(48):19444-19449.

145. Plasschaert LW, Zilionis R, Choo-Wing R, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature.* 2018;560(7718):377-381.

146. Doupe DP, Alcolea MP, Roshan A, et al. A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science.* 2012;337(6098):1091-1093.

147. DeWard AD, Cramer J, Lagasse E. Cellular heterogeneity in the mouse esophagus implicates the presence of a nonquiescent epithelial stem cell population. *Cell Rep.* 2014;9(2):701-711.

148. Giroux V, Lento AA, Islam M, et al. Long-lived keratin 15+ esophageal progenitor cells contribute to homeostasis and regeneration. *J Clin Invest.* 2017;127(6):2378-2391.

149. Zhang Y, Yang Y, Jiang M, et al. 3D Modeling of Esophageal Development using Human PSC-Derived Basal Progenitors Reveals a Critical Role for Notch Signaling. *Cell Stem Cell.* 2018;23(4):516-529 e515.

150. Bingle L, Cross SS, High AS, et al. WFDC2 (HE4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. *Respir Res.* 2006;7:61.

151. Li J, Chen H, Mariani A, et al. HE4 (WFDC2) Promotes Tumor Growth in Endometrial Cancer Cell Lines. *Int J Mol Sci.* 2013;14(3):6026-6043.

152. Jones KB, Furukawa S, Marangoni P, et al. Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell.* 2019;24(1):183-192 e188.

153. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet.* 2005;37(4):382-390.

154. Lord RV, Brabender J, Wickramasinghe K, et al. Increased CDX2 and decreased PITX1 homeobox gene expression in Barrett's esophagus and Barrett's-associated adenocarcinoma. *Surgery.* 2005;138(5):924-931.

155. Chen H, Zhang M, Zhang W, et al. Downregulation of BarH-like homeobox 2 promotes cell proliferation, migration and aerobic glycolysis through Wnt/beta-catenin signaling, and predicts a poor prognosis in non-small cell lung carcinoma. *Thorac Cancer.* 2018;9(3):390-399.

156. Olson LE, Zhang J, Taylor H, Rose DW, Rosenfeld MG. Barx2 functions through distinct corepressor classes to regulate hair follicle remodeling. *Proc Natl Acad Sci U S A.* 2005;102(10):3708-3713.

157. Araya J, Cambier S, Markovics JA, et al. Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. *J Clin Invest.* 2007;117(11):3551-3562.

158. Warnecke-Eberz U, Metzger R, Holscher AH, Drebber U, Bollschweiler E. Diagnostic marker signature for esophageal cancer from transcriptome analysis. *Tumour Biol.* 2016;37(5):6349-6358.

159. Leclerc EA, Huchenq A, Kezic S, Serre G, Jonca N. Mice deficient for the epidermal dermokine beta and gamma isoforms display transient cornification defects. *J Cell Sci.* 2014;127(Pt 13):2862-2872.

160. Chen HL, Chiang PC, Lo CH, et al. Galectin-7 Regulates Keratinocyte Proliferation and Differentiation through JNK-miR-203-p63 Signaling. *J Invest Dermatol.* 2016;136(1):182-191.

161. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006;126(4):663-676.

162. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature.* 2010;463(7284):1035-1041.

163. Xiao D, Liu X, Zhang M, et al. Direct reprogramming of fibroblasts into neural stem cells by single non-neural progenitor transcription factor Ptf1a. *Nat Commun.* 2018;9(1):2865.

164. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature.* 2008;455(7213):627-632.

165. Banga A, Akinci E, Greder LV, Dutton JR, Slack JM. In vivo reprogramming of Sox9+ cells in the liver to insulin-secreting ducts. *Proc Natl Acad Sci U S A.* 2012;109(38):15336-15341.

166. Yang Y, Akinci E, Dutton JR, Banga A, Slack JM. Stage specific reprogramming of mouse embryo liver cells to a beta cell-like phenotype. *Mech Dev.* 2013;130(11-12):602-612.

167. Califano A, Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer.* 2017;17(2):116-130.

168. Aytes A, Giacobbe A, Mitrofanova A, et al. NSD2 is a conserved driver of metastatic prostate cancer progression. *Nat Commun.* 2018;9(1):5201.

169. Talos F, Mitrofanova A, Bergren SK, Califano A, Shen MM. A computational systems approach identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. *Nat Commun.* 2017;8:14662.

170. Carro MS, Lim WK, Alvarez MJ, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature.* 2010;463(7279):318-325.

171. Ning W, Li CJ, Kaminski N, et al. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A.* 2004;101(41):14895-14900.

172. Bai J, Hu S. Transcriptome network analysis reveals potential candidate genes for squamous lung cancer. *Int J Mol Med.* 2012;29(1):95-101.

173. Yeo J, Morales DA, Chen T, et al. RNAseq analysis of bronchial epithelial cells to identify COPD-associated genes and SNPs. *BMC Pulm Med.* 2018;18(1):42.

174. Didon L, Roos AB, Elmberger GP, Gonzalez FJ, Nord M. Lung-specific inactivation of CCAAT/enhancer binding protein alpha causes a pathological pattern characteristic of COPD. *Eur Respir J.* 2010;35(1):186-197.

175. Pan J, Qin Y, Zhang M. HPIP promotes non-small cell lung cancer cell proliferation, migration and invasion through regulation of the Sonic hedgehog signaling pathway. *Biomed Pharmacother.* 2016;77:176-181.

176. Kwon E, Todorova K, Wang J, et al. The RNA-binding protein YBX1 regulates epidermal progenitors at a posttranscriptional level. *Nat Commun.* 2018;9(1):1734.

177. Martin M, Hua L, Wang B, et al. Novel Serine 176 Phosphorylation of YBX1 Activates NF-kappaB in Colon Cancer. *J Biol Chem.* 2017;292(8):3433-3444.

178. Singh V, Singh LC, Singh AP, et al. Status of epigenetic chromatin modification enzymes and esophageal squamous cell carcinoma risk in northeast Indian population. *Am J Cancer Res.* 2015;5(3):979-999.

179. Dong Z, Liu S, Zhou C, et al. Overexpression of Id-1 is associated with tumor angiogenesis and poor clinical outcome in oral squamous cell carcinoma. *Oral Oncol.* 2010;46(3):154-157.

180. Pillai S, Rizwani W, Li X, et al. ID1 facilitates the growth and metastasis of non-small cell lung cancer in response to nicotinic acetylcholine receptor and epidermal growth factor receptor signaling. *Mol Cell Biol.* 2011;31(14):3052-3067.

181. Dixit A, Parnas O, Li B, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell.* 2016;167(7):1853-1866 e1817.

182. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods.* 2017;14(3):297-301.